



THE ROLE OF METAMEMORY IN EYEWITNESS TESTIMONY PERFORMANCE

By: Renan Benigno Saraiva

July 2019

This thesis is submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the University of Portsmouth.

Supervisors:

Professor Lorraine Hope

Dr. James Ost

Dr. Robert Horselenberg

Professor Peter van Koppen

This research was supported by a fellowship award from the Erasmus Mundus Joint Doctorate Program The House of Legal Psychology (EMJD-LP) with Framework Partnership Agreement (FPA) 2013-0036 and Specific Grant Agreement (SGA) 2013-0036.



The House of
Legal Psychology

GENERAL ABSTRACT

Estimating eyewitness memory accuracy is crucial in forensic settings, given the need for efficient investigations and the negative consequences of erroneous testimony. In this thesis, the overarching goal of the research was to test the utility of metamemory assessments as postdictors of eyewitness performance. Metamemory research is essential for a comprehensive understanding of how people use and perceive their own memory, but it has not yet been thoroughly explored in eyewitness settings. In the initial experiments (**Experiments 1 and 2**), the relationship between self-reports of memory ability and eyewitness identification performance was examined. Specifically, we tested how self-ratings of memory ability and strategies endorsement relate to eyewitness identification accuracy, confidence and over/underconfidence. These experiments provided initial evidence that some metamemory factors are important indicators of eyewitness identification accuracy and confidence, although only assessments of global memory self-efficacy were adopted. Aiming to improve metamemory assessments in eyewitness contexts, a measure tailored specifically to eyewitness testimony settings was developed and tested in **Experiment 3**) In **Experiment 4**, the predictive value of general and eyewitness-specific metamemory assessments (EMS) on identification performance for biased and unbiased lineups was investigated. We were specifically interested in possible differential effects between biased and unbiased lineups because other postdictors of identification performance have less diagnostic value in biased lineups. The results of **Experiment 4** showed that eyewitness-specific metamemory factors are predictive of identification accuracy for both biased and unbiased lineups. In **Experiment 5** an innovative repeated-trials eyewitness identification paradigm was used to examine the relation between metamemory assessments and identification performance, obtaining further evidence for a relation between eyewitness-specific metamemory factors and identification

performance. In the final experiment (**Experiment 6**), the focus is shifted from identification paradigms to an eyewitness free recall paradigm, aiming to elucidate the relation between metamemory and disclosure of information. We summarize the main findings of this novel line of research in the **General Discussion**, presenting the challenges and prospects facing future eyewitness metamemory research.

TABLE OF CONTENTS

CHAPTER 1

General Introduction: Examining the Intersection Between Metamemory and Eyewitness Testimony Research	Error! Bookmark not defined.
Theoretical Frameworks of Metamemory	Error! Bookmark not defined.
Judgments of Confidence and Memory Self-Efficacy	Error! Bookmark not defined.
Metamemory in Eyewitness Testimony Contexts	Error! Bookmark not defined.
Eyewitness Confidence-Accuracy Relationship	Error! Bookmark not defined.
Outline of the Current Thesis	Error! Bookmark not defined.

CHAPTER 2

Estimating Eyewitness Identification Performance Using Metamemory Measures and Memory Tests	Error! Bookmark not defined.
Metamemory assessments and eyewitness identification performance	Error! Bookmark not defined.
Memory tests and eyewitness identification performance	Error! Bookmark not defined.
Experiment 1	Error! Bookmark not defined.
Experiment 2	Error! Bookmark not defined.
General Discussion	Error! Bookmark not defined.

CHAPTER 3

Development and validation of the Eyewitness Metamemory Scale ..	Error! Bookmark not defined.
Materials and Methods	Error! Bookmark not defined.
Results	Error! Bookmark not defined.
Discussion	Error! Bookmark not defined.

CHAPTER 4

Eyewitness Metamemory Predicts Identification Performance in Biased and Unbiased Lineups	Error! Bookmark not defined.
Method	Error! Bookmark not defined.
Results	Error! Bookmark not defined.
Discussion	Error! Bookmark not defined.

CHAPTER 5

Using General and Eyewitness-Specific Metamemory Assessments to Estimate Performance in Multiple Identifications	Error! Bookmark not defined.
Method	Error! Bookmark not defined.
Results	Error! Bookmark not defined.

Discussion.....	Error! Bookmark not defined.
CHAPTER 6	
Using Metamemory Measures and Memory Tests to Estimate Eyewitness Free Recall Performance.....	Error! Bookmark not defined.
Method.....	Error! Bookmark not defined.
Results	Error! Bookmark not defined.
Discussion.....	Error! Bookmark not defined.
CHAPTER 7	
General Discussion	Error! Bookmark not defined.
Summary of the Findings	Error! Bookmark not defined.
Metamemory and Identification Accuracy	Error! Bookmark not defined.
Metamemory and Identification Confidence-Accuracy Relation....	Error! Bookmark not defined.
Metamemory and Reporting of Information	Error! Bookmark not defined.
Objective Tests of Memory Performance.....	Error! Bookmark not defined.
Methodological Limitations	Error! Bookmark not defined.
Implications of this Research	Error! Bookmark not defined.
Future Research Directions	Error! Bookmark not defined.
Conclusion	Error! Bookmark not defined.
REFERENCES	Error! Bookmark not defined.
APPENDICES	Error! Bookmark not defined.

LIST OF FIGURES AND TABLES

FIGURES

<i>Figure 2.1.</i> Posterior distributions of coefficients of metamemory factors as predictors of identification accuracy in Experiment 1 for Choosers (A) and Nonchoosers (B). The shaded area in the distributions represent 95% credible intervals.	33
<i>Figure 2.2</i> Posterior distributions of coefficients of metamemory factors as predictors of identification confidence in Experiment 1 for Choosers (A) and Nonchoosers (B). The shaded area in the distributions represent 95% credible intervals.	34
<i>Figure 2.3.</i> Calibration curve for Choosers identification performance by higher and low scorers in MMQ-Strategy (Experiment 1). The diagonal dotted grey line represents perfect calibration.	36
<i>Figure 2.4.</i> Inferential confidence intervals of calibration statistics for high and low scorers in each metamemory measure in Experiment 1.	36
<i>Figure 2.5.</i> Posterior distributions of coefficients of metamemory factors as predictors of identification accuracy in Experiment 2 for Choosers (A) and Nonchoosers (B). The shaded area in the distributions represent 95% credible intervals.	40
<i>Figure 2.6.</i> Posterior distributions of coefficients of metamemory factors as predictors of identification confidence in Experiment 2 for Choosers (A) and Nonchoosers (B). The shaded area in the distributions represent 95% credible intervals.	40
<i>Figure 2.7.</i> Calibration curve for Choosers identification performance by higher and low scorers in MMQ-Strategy (Experiment 2). The diagonal dotted grey line represents perfect calibration.	41
<i>Figure 2.8.</i> Inferential confidence intervals of calibration statistics for high and low scorers in each metamemory measure in Experiment 2.	42
<i>Figure 3.1.</i> Model specification and model fit indices for the three-factor solution (A), revised three-factor solution (B) and two-factor solution (C) of the Eyewitness Metamemory Scale.....	65
<i>Figure 4.1.</i> Predicted probabilities of identification accuracy among choosers as a function of EMS-Discontentment for biased and unbiased lineups. The shaded polygon represents 95% confidence intervals.	84
<i>Figure 4.2.</i> Confidence-accuracy calibration curves comparing choosers and nonchoosers (A) and biased vs unbiased lineups (B). The dotted diagonal grey line represents perfect calibration.	86

<i>Figure 4.3.</i> Inferential confidence intervals of calibration statistics for high and low scorers in each metamemory measure.	87
<i>Figure 4.4.</i> Calibration curves comparing low and high scorers in the EMS-Contentment, EMS-Discontentment, EMS-Strategy and SSMQ metamemory factors.	88
<i>Figure 5.1.</i> Multilevel regression plots of predicted probabilities for model outcomes as a function of EMS-Contentment. Regression lines are predicted relationships and shaded polygons represent 95% confidence intervals.	106
<i>Figure 6.1.</i> Calibration curves of eyewitness free recall and all memory tests. Error bars are 95% confidence intervals.	128
<i>Figure 6.2.</i> Inferential confidence intervals of calibration statistics for high and low scorers in each metamemory measure.	129

TABLES

<i>Table 2.1.</i> Logistic Regressions Testing Metamemory Measures as Predictors of Eyewitness Identification Accuracy (Experiment 1).....	33
<i>Table 2.2.</i> Multiple Regressions Testing Metamemory Measures as Predictors of Eyewitness Confidence (Experiment 1).....	34
<i>Table 2.3.</i> Logistic and Multiple Regression Analyses Examining Performance in Memory Tests as Predictors of Eyewitness Identification Accuracy and Confidence (Experiment 1)	37
<i>Table 2.4.</i> Logistic Regressions Testing Metamemory Measures as Predictors of Eyewitness Identification Accuracy (Experiment 2).....	39
<i>Table 2.5.</i> Multiple Regressions Testing Metamemory Measures as Predictors of Eyewitness Confidence (Experiment 2).....	39
<i>Table 3.1.</i> Item-total correlations, communalities and pattern matrix for the Eyewitness Metamemory Scale.....	63
<i>Table 3.2.</i> Means, standard deviations, and correlations with confidence intervals of the EMS factors and other metamemory measures	66
<i>Table 4.1.</i> Means, Standard Deviations, and Correlations of the Metamemory Factors	82
<i>Table 4.2.</i> Logistic Regression Models of Metamemory Factors as Predictors of Identification Accuracy Among Choosers	83
<i>Table 4.3.</i> Logistic Regression Models of Metamemory Factors as Predictors of Identification Accuracy Among Nonchoosers	83
<i>Table 4.4.</i> Calibration Statistics for Choosers and Nonchoosers Across Biased and Unbiased Conditions	85
<i>Table 5.1.</i> Main descriptive statistics for identification performance scores	104
<i>Table 5.2.</i> Means, standard deviations, and correlations of the metamemory measures ..	104

<i>Table 5.3.</i> Raw p-values and adjusted p-values using Benjamini-Hochberg correction for all regression models	105
<i>Table 5.4.</i> Logistic multilevel models testing metamemory scores as predictors of eyewitness identification performance	105
<i>Table 5.5.</i> Multiple regressions testing metamemory scores as predictors of calibration and over/underconfidence	107
<i>Table 6.1.</i> Means and standard deviations of the main memory performance variables ..	124
<i>Table 6.2.</i> Bayesian Multiple linear regression models including metamemory scales as predictors of eyewitness free recall performance	125
<i>Table 6.3.</i> Regression models using performance in the different memory tests as predictors of eyewitness free recall performance	126
<i>Table 6.4.</i> Calibration statistics for each memory test with inferential confidence intervals (ICI)	128

ABBREVIATIONS

ANOVA: Analysis of variance

CA: Confidence-accuracy

EMS: Eyewintess Metamemory Scale

ICC: Intra-class correlation

MMQ: Multifactorial Metamemory Questionnaire

SSMQ: Squire Subjective Memory Questionnaire

DECLARATION

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Word count: 41.865

ACKNOWLEDGEMENTS

This work and my experience with it would not have had the same spirit without the invaluable academic and human support I received in the past three years. Foremost, I would like to express my gratitude to the founders and all members of the House of Legal Psychology. It has been a great privilege to be part of such a diverse, enthusiastic and inspiring programme.

It is an honour for me to have worked alongside so many brilliant collaborators. I must first pay tribute to the high quality of supervision given by *Lorraine Hope*. Without your patient guidance, enthusiastic encouragement and useful critiques this work would hardly have been completed. Above all, I can remember many times when I felt discouraged and unsure of the intrinsic worth of my own work, but you never failed to offer timely advice and encouragement when it was most needed. I am also deeply grateful to my co-supervisors, *Peter van Koppen* and *Robert Horselenberg*. Your guidance and deep insight into real-world issues brought to my attention important gaps in my knowledge and holes in my logic. I am also in debt with key contributors to this work. *James Sauer* and *Melanie Sauerland*, your supportive and knowledgeable feedback were invaluable and it has been a great pleasure to work with you. Finally, I would like to extend my heartfelt thanks to *James Ost*. You were an incredibly inspirational person and a role model to be followed.

I would like to express my deep gratitude to my dear friends at home (Brazil) and all across the world. Your support and companionship fill me with determination and aspiration to be better.

Dear Giulia, it is amazing how much joy and fulfilment you add to my life. Thank you for always being by my side.

To my parents Carlos and Águeda, without your love and support I would never have gotten this far.

DISSEMINATION

Conference presentations

Saraiva, R.B., Van Boeijen, I. M., Hope, L, Horselenberg, R., Sauerland, M., & Van Koppen, P.J. (2019, July). Development and validation of the Eyewitness Metamemory Scale. Poster presentation at the annual conference of the European Association of Psychology and Law (EAPL), Santiago de Composterla, Spain.

Saraiva, R.B., Van Boeijen, I. M., Hope, L., Horselenberg, R., & Van Koppen, P.J. (2019, July). Using general and eyewitness-specific metamemory assessments to estimate performance in multiple identifications. Oral presentation at the annual conference of the European Association of Psychology and Law (EAPL), Santiago de Composterla, Spain.

Saraiva, R.B., Van Boeijen, I. M., Hope, L, Horselenberg, R., Sauerland, M., & Van Koppen, P.J. (2019, July). Eyewitness metamemory predicts identification performance in biased and unbiased lineups. Oral presentation at the annual conference of the European Association of Psychology and Law (EAPL), Santiago de Composterla, Spain.

Saraiva, R.B., Hope, L, Horselenberg, R., Ost, J., Sauer., J & Van Koppen, P.J. (2018, July). Estimating eyewitness identification performance using metamemory measures and memory tests. Oral presentation at the annual conference of the European Association of Psychology and Law (EAPL), Turku, Finland.

Saraiva, R.B., Hope, L, Horselenberg, R., Ost, J., Sauer., J & Van Koppen, P.J. (2018, July). Estimating eyewitness free recall performance using metamemory measures and

memory tests. Poster presentation at the annual conference of the International Investigative Interviewing Research Group (iIIRG), Porto, Portugal.

Publications

Saraiva, R. B., Van Boeijen, I. M., Hope, L., Horselenberg, R., Sauerland, M., & Van Koppen, P (in press). Development and validation of the Eyewitness Metamemory Scale.

Applied Cognitive Psychology.

Saraiva, R.B., Hope, L, Horselenberg, R., Van Boeijen, I. M., Sauer., J., & Van Koppen, P.

Using general and eyewitness-specific metamemory assessments to estimate performance in multiple identifications. *Journal of Applied Memory and Cognition*

Saraiva, R.B., Hope, L., Horselenberg, R., Ost, J., Sauer, J., & Van Koppen, P.J. (under

review). Estimating eyewitness free recall performance using metamemory measures and memory tests. *Memory*

Saraiva, R.B., Hope, L, Horselenberg, R., Ost, J., Sauer., J., & Van Koppen, P. (under

review). Estimating eyewitness identification performance using metamemory measures and memory tests. *Law and Human Behavior*.

Saraiva, R.B., Van Boeijen, I.M., Hope, L., Horselenberg, R., Sauerland, M., & Van Koppen,

P.J. (under review). Eyewitness metamemory predicts identification performance in biased and unbiased lineups. *Legal and Criminological Psychology*

CHAPTER 1

General Introduction:

Examining The Intersection Between Metamemory And Eyewitness

Testimony Research

Eyewitnesses are crucial in the criminal justice system, providing evidence that may be used to guide investigations and to identify perpetrators. Since the initial formulation of eyewitness testimony research (Munsterberg, 1908), this maturing field has established that eyewitness memory is not a permanent record of perceived events and may be tainted by a number of different factors. Witnesses' recollections of important facts not only deteriorate over time but can also be negatively distorted by new information introduced after the original experience. Alarmingly, decades of research confirms that individuals may be led to believe that they experienced events that never actually took place, such as getting lost on a mall (Loftus & Pickrell, 1995), spilling punch at a wedding (Hyman & Billings, 1998), or even of committing a crime (Kassin, 2008; Kassin & Kiechel, 1996). In less extreme circumstances, eyewitnesses may incorporate false information into their accounts after communicating with other witnesses (Ito et al., 2018), or in response to suggestive interviewing practices (Eisen, Quas, & Goodman, 2001). Such malleability of memory does not imply that eyewitness evidence is inherently unreliable, but it emphasizes how eyewitness accounts can be contaminated (Wixted, Mickes, & Fisher, 2018).

Similarly, the processes by which we monitor our memory performance are also not perfectly objective. In forensic settings, eyewitness confidence in identification procedures can be inflated by external influences, such as confirmatory feedback given by legal officers (Douglass & Steblay, 2006; Wells & Bradfield, 1998), or by the use of biased lineups (Charman, Wells, & Joy, 2011). Even in the absence of external influences, confidence judgements may be undermined when eyewitnesses have a lack of insight into their own memory capacity (Perfect, 2004). Such failures in recollections and memory monitoring processes can have severe consequences in the legal justice system. Namely, confidence statements may be less diagnostic of accuracy if eyewitnesses are not realistic about their self-perceived memory ability (Olsson & Juslin, 1999), or eyewitnesses may underreport

critical aspects of what they remember due to self-distrust in memory ability (Goldsmith, Pansky, & Koriati, 2014; Weber & Brewer, 2008). Surprisingly, some of the potential issues concerning self-assessments of memory capacity and eyewitness testimony have not been thoroughly investigated.

The goal of the programme of research conducted for this thesis was to address this gap and investigate the relation between memory self-assessments and different aspects of eyewitness testimony performance. Specifically, I explored three overarching research goals in detail across this programme of work. The first goal was to determine the relationship between self-reported metamemory and objective memory accuracy in different eyewitness paradigms. The second goal was to examine the relation between metamemory and expressions of confidence in eyewitness settings. Finally, the third goal was to investigate the relation between metamemory and the confidence-accuracy relationship in eyewitness performance. The experiments designed to achieve these goals were formalized using theoretical frameworks derived from metamemory research. This chapter presents a definition and general overview of metamemory as a psychological construct and introduces the many potential implications of metamemory frameworks when planning and conducting empirical studies on eyewitness testimony.

Theoretical Frameworks of Metamemory

We all need to make judgements about the likely accuracy of our memories from time to time. An interviewer relying on his or her notes must make this judgement, as must a patient reporting their medical history, as must an eyewitness sworn to tell the truth in Court. The ability to examine one's own memory performance is a critical feature of normal metacognitive functioning. Metamemory, an aspect of metacognition, is a latent construct that was formulated from the early developmental psychology literature to describe the

knowledge, perceptions, and beliefs individuals have about their own memory and the memory system in more general terms (Hultsch, Hertzog, Dixon, & Davidson, 1988). These beliefs and judgments greatly influence behaviour because they are used to monitor and control how we deploy cognitive resources (Koriat, Ma'ayan, & Nussinson, 2006). For example, if a student believes some important material was not properly learned, she or he may spend more time studying that material or employ strategies to better memorize its content. Thus, metamnemonic judgments indicate how well target items are either available or accessible in memory. For instance, an individual may be able to retrieve an online-password now, but metamnemonic judgments assess whether the same password will be retrievable at some future date. Similarly, at test, an individual may decide that even though she or he may not know the password now, she or he has a high 'feeling of knowing' for the item, so the individual chooses to spend more time trying to retrieve the password.

Both memory monitoring and control play an important role in a variety of everyday life situations. For instance, depending on monitoring results, not-mastered materials can be further studied, a more effective learning strategy can be adopted, or external cues can be used to improve remembering. The efficacy of this metacognitive system requires a feedback system between monitoring and control mechanisms, which is provided by metamemory judgements. Several types of metamemory judgements have been investigated including reality and source monitoring (Johnson, Hashtroudi, & Lindsay, 1993), retrospective confidence judgments (Brewer & Sampaio, 2012), "don't know" judgments (Tobias & Everson, 2009), hindsight judgments (Koriat & Bjork, 2005), and judgments of subjective experience (Koriat et al., 2006; Schwartz, Benjamin, & Bjork, 1997).

Following Nelson and Narens (1990), metamemory judgments can be divided into those made during acquisition of knowledge (judgments of learning; JOL) and judgments made at the time of retrieval (feelings of knowing; FOK). Judgments of learning can be

defined as prospective confidence judgments of encoding efficiency made after exposure to an item but prior to a recall test. In forensic settings, judgements of learning can be observed when eyewitnesses of a crime are asked whether they would be able to recognize the perpetrator or produce a facial composite. In this scenario, the eyewitness needs to reflect on their memory ability, their internal state, and on the encoding conditions when deciding whether a recognition would be possible. In contrast, feeling of knowing can be defined as a retrospective confidence judgement of encoding efficiency made during the time of retrieval. After making a lineup identification, for example, eyewitnesses may express how confident they are in their decision. Confidence judgments are one commonly used method for determining an individual's belief that the information retrieved from memory is accurate.

There have been some attempts to develop general theories of memory confidence, such as direct access theories (Schwartz, 1994), familiarity based theories (Metcalf, 1996; Metcalf, 2000), and accessibility theories (Dunlosky, 2004). A common assumption across those theories is that the processes underlying confidence judgements fall into two broad categories: target-based sources and cue-based sources. That is, confidence judgments are based on many different characteristics of memory retrieval processes, such as memory vividness and completeness (Talarico & Rubin, 2003), response latency (Brewer, Caon, Todd, & Weber, 2006), and the quantity and intensity of retrieved information (Koriat, 1993). Target-based sources include the ease of stimuli processing, the amount of information encoded, or interference from distractors. In direct-access theories, for example, there is an emphasis on target-based information so that people monitor the specific memory representation of an item when making metamnemonic judgments (Schwartz, 1994). Cue based sources include information that is not intrinsic to the target, such as familiarity with the stimuli, domain knowledge, and social desirability (Metcalf, 1996; Metcalf, 2000). Cued-based sources play an important role in familiarity based-theories and accessibility

theories, positing that people draw on information other than the specific representation to form a confidence judgment.

One influential theoretical framework proposed by (Koriat, 2000) extends on the processes leading to metacognitive judgements just outlined. In this framework, metacognitive judgements can be based on information (or theory-based) and experience (or affect-based). That is, monitoring memory processes can be based on an explicitly inferential process, or on a sheer subjective feeling. Consider, for example, an eyewitness who fails to recall a specific fact of a crime or a characteristic of the perpetrator. The witness may still be able to make an educated guess about the plausibility that the solicited information will be subsequently recalled or recognized. Such judgment may be based on domain-specific memories and beliefs (Nelson, Gerler, & Narens, 1984), and can take the form “there is little chance that I would know the answer” or “I ought to know the answer” (Costermans, Lories, & Ansay, 1992).

Judgments of Confidence and Memory Self-Efficacy

The degree of confidence that an individual expresses in a memory plays a critical role in how an outsider evaluates the verity of that memory (Wells, Olson, & Charman, 2002). In addition, in many applied areas such as jury decisions and medical diagnosis, confidence in one’s judgements determines the likelihood of translating these judgements to action. But the functional value of confidence depends on whether people are generally accurate in monitoring their memories or knowledge. Results of many studies have demonstrated positive correlations across items between subjective confidence and objective accuracy in knowledge tests, suggesting that people have good insights into the relative accuracy of their knowledge. This has been found to be the case across a variety of metacognitive judgments. Judgments of learning made about different items during study are moderately predictive of the relative future recall or recognition of these items (Narens,

Nelson, & Scheck, 2008). Similarly, feeling of knowing judgments following a recall failure are predictive of the likelihood of recalling the target at some later time (Koriat, 1995).

Finally, confidence judgments in general knowledge answers are generally diagnostic of the answers correctness (Koriat, 2008). Nevertheless, disconnection between subjective and objective indexes of knowing have been observed in some situations to the extent that metacognitive judgments were not diagnostic of actual memory performance (Benjamin, Bjork, & Schwartz, 1998; Besken & Mulligan, 2013).

A specific dimension of metamemory that is relevant to everyday memory functioning is memory self-efficacy (MSE), defined as the self-evaluation of one's general competence and ability across many different memory domains and tasks (Berry, 1999; Hertzog & Dixon, 1994). Hertzog and Dixon (1994) propose a hierarchical structure for this construct including global beliefs about memory ability (e.g., "I have a good memory") and situational beliefs ("I can remember this address, so I will go there without looking it up"). Unlike situational beliefs, global MSE beliefs are enduring and have been constructed on the basis of previous experiences and implicit theories and schemas about memory (e.g., perceived effects of ageing on memory, perceived memory stability across domains; (Cavanaugh & Green, 1990; Hertzog, McGuire, & Lineweaver, 1998). Global MSE beliefs may also be defined in relation to an individual's appraisal of his or her ability in a specific memory domain such as semantic memory, memory for faces, or episodic memory (Hertzog & Dixon, 1994; Hertzog, Park, Morrell, & Martin, 2000). Some evidence points to a positive relation between memory self-efficacy and memory performance in different tasks (Seeman, McAvay, Merrill, Albert, & Rodin, 1996; Valentijn et al., 2006). In the face matching literature, some results show moderate to large correlations between self-reported face perception ability and performance in face matching tests (Gray, Bird, & Cook, 2017; Ventura, Livingston, & Shah, 2018). However, the outcome of studies focusing specifically

on face recognition show that individuals have limited insight into their ability to recognize unfamiliar faces (Bindemann, Attard, & Johnston, 2014; Bobak, Mileva, & Hancock, 2018).

Despite an emerging literature on the relation between memory self-efficacy and objective memory performance, very little of this research has focused on eyewitness testimony settings. One of the goals of the current programme of work was to address this gap. Across all experiments in the current thesis we examine the relation between memory self-efficacy and eyewitness testimony performance. Specifically, we aimed to examine the utility of memory self-efficacy as a predictor of accuracy, confidence and confidence-accuracy relation in eyewitness identification and free recall tasks. From a theoretical perspective, we aimed to elucidate how insightful individuals are about their own memory performance in forensic contexts. From a practical perspective, we examined whether assessments of memory-efficacy may be used to better distinguish accurate from inaccurate eyewitness identifications or accounts.

Metamemory in Eyewitness Testimony Contexts

Metamemory processes are implicated at various stages of memory formation, and those operating during remembering are particularly crucial in determining memory accuracy and error. These processes have several functions that are notably relevant in eyewitness testimony contexts, such as: specifying the origin of mental experience (Mitchell & Johnson, 2000), avoiding suggestibility effects and memory contamination by attributing them to their proper source (Mitchell, Johnson, & Mather, 2003), adopting appropriate strategies to enhance memory for the task at hand (Bjorklund, Dukes, & Brown, 2008), formulating a comprehensive and informative account of a past event (Koriat, 1999), monitoring the accuracy of remembered information (Kelemen, 2000), and regulating the reporting of information according to the incentive for accuracy (Koriat & Goldsmith, 1996). Despite a

clear overlap between metamemory processes and eyewitness testimony settings, there is still a considerable gap in the literature concerning the intersections between metamemory and eyewitness performance. Across this programme of research, the thesis I aim to advance is that in order to improve or better estimate memory accuracy, it is crucial to understand how people monitor and evaluate their own memory cognitive processes.

Metacognitive monitoring has been studied in many different contexts, such as education, language learning, problem-solving, and teaching, offering an integrated theoretical framework to explain how people evaluate and judge their own memories (Hacker, Dunlosky, & Graesser, 2009). Trace access theory, for example, postulates that both confidence and accuracy are dependent on memory trace strength. Stronger memory traces, or information with a higher number of associated stimuli, is more likely to be remembered and receive higher confidence ratings (Burke, MacKay, Worthley, & Wade, 1991; Busey, Tunnicliff, Loftus, & Loftus, 2000; Clark, 1997). Findings regarding the optimality hypothesis reported by Deffenbacher (1980) are consistent with trace access theory, so that witnesses with better encoding conditions are more capable of making correct identifications and accurate judgments about their performance. However, studies focusing on understanding why optimal conditions increase the confidence-accuracy (CA) relationship seems sparse. Better viewing conditions provide witnesses with better memory traces, leading to higher accuracy rates. But why do optimal conditions lead to higher CA relationship is a more challenging question. Leippe and Eisenstadt (2014) argue that better access to memory evaluation may allow witnesses to correctly reduce their confidence when they are inaccurate. They also propose that witnesses with an overall weak memory would have weaker memory traces despite optimal conditions and would use other sources of information - including their weak memory traces - to make confidence judgements. In both cases

inaccurate information would be more likely to be followed by low confidence decisions, increasing the CA relationship.

Despite its contribution, trace access theory ignores many important aspects of memory confidence judgements. Information derived from a memory trace is just one of many sources of information that influence witnesses metamemory and beliefs. Witnesses making a self-evaluation to gauge confidence can take into account many intrinsic, heuristic and self-credibility cues, especially when the memory trace is weak (Leippe & Eisenstadt, 2014). Intrinsic cues include self-reflections on memory processes, from which people identify information that they learned to associate with accurate or inaccurate memory. For example, Kebbell, Wagstaff and Covey (1996), found that witnesses showed less confidence in their correct answers to hard questions when compared to their correct answers to easy questions, possibly because questions difficulty is used as a heuristic to judge confidence. Furthermore, witnesses usually rely on recognition speed to make inferences about their confidence, showing greater confidence when they are able to make rapid judgements (Robinson, Johnson, & Herndon, 1997).

Heuristic cues comprise one's beliefs about external factors that can help or impair memory encoding and retrieval. It may be the case, for example, that witnesses feel more confident when they put more effort on recall tasks, even when this does not lead to changes in accuracy (Shaw & Zerr, 2003). People also usually believe that their memory will be stronger with prolonged exposure duration, consequently showing higher confidence for stimuli they have been exposed to longer (Memon, Hope, & Bull, 2003). Finally, self-credibility cues derive from people's self-evaluation of their overall memory performance. Some could overestimate their ability to recall events, being overconfident on identification and recall tasks; others may underestimate their memory ability and show underconfidence (Leippe, Eisenstadt, Rauch, & Stambush, 2006; Olsson & Juslin, 1999). One central goal in

the programme of work comprising this thesis was to examine whether self-ratings of memory performance relate to eyewitness performance. If confidence statements and memory traces are partially based on heuristic cues, it can be expected that self-ratings of memory capacity are predictive of performance in eyewitness memory tasks. That is, individual's self-evaluations of their own memory performance may be related to their accuracy, confidence and over/underconfidence in identification and free recall settings, although this hypothesis has not yet been thoroughly explored.

Eyewitness Confidence-Accuracy Relationship

Criminal justice systems usually rely on eyewitnesses' confidence when evaluating the likely guilty of a defendant or a suspect. The U.S. Supreme Court, for instance, endorses eyewitnesses' confidence as one criterion for assessing their accuracy (Neil v. Biggers, 1972). Literature also shows that eyewitnesses' confidence influences decisions made by police officers, lawyers and jurors (e.g., Bradfield & Wells, 2000). For example, jurors are more likely to believe testimony when it is provided by a confident, rather than uncertain, witness (Brewer & Burke, 2002). So, although confidence measures can be central to criminal investigations, they may lead to undesirable consequences for justice if eyewitnesses are overconfident or underconfident. Overconfident witnesses may lead investigators to follow incorrect leads, or - more severely - incriminate an innocent person (The Innocence Project, 2019). Underconfident witnesses may withhold relevant information, or be insufficiently confident even when a perpetrator is correctly identified.

Results from most of the earliest studies on witness confidence-accuracy relationship (CA) showed that confidence and accuracy were often poorly correlated (e.g., Cutler & Penrod, 1989; Deffenbacher, 1980; Leippe, 1980). A few years later Bothwell and Deffenbacher (1987) conducted a meta-analysis of 35 staged-event studies, revealing an

estimated effect size of $r = 0.25$ ($d = .52$) with a 95% CI [.08 to .042], wherein 52% of the variation in r was attributed to sampling errors, still recommending caution when using confidence to predict identification accuracy in actual cases. At the turn of the millennium, the majority of experts in eyewitness memory interviewed by Kassin, Tubb, Hosch, and Memon (2001) agreed that confidence is not a good marker of accuracy. However, new approaches to statistical inference in the field of eyewitness testimony revealed two important aspects that challenged the previously negative conclusions about CA relationship: the role of calibration approaches and likely impact of moderator variables.

Calibration approaches draw on the principle that there are multiple ways of calculating the relationship between confidence and accuracy. The most conventional method, for example, simply compares the accuracy of individuals who differ in confidence. This calculation is basically a point-biserial correlation between confidence and accuracy across the witnesses, usually used in studies where many witnesses testify about a single event. This approach is suitable for answering the question of whether a more confident witness is also likely to be a more accurate witness, or vice versa. However, some researchers (e.g., Juslin, Olsson, & Winman, 1996; Olsson, 2000; Weber & Brewer, 2003) have argued that the point-biserial correlation is not the most informative measure of the CA relationship, because it does not express the extent to which confidence corresponds to actual probability of accuracy. An alternative approach is the use of calibration, which aims to verify if a witness is more accurate about the information for which they show the greatest confidence. In this approach the confidence-accuracy relationship is computed across every level of confidence, typically in 10% increments (e.g., a scale that ranges from 0% - “not at all certain”, to 100% - “totally certain”). A perfect calibration occurs when the group's confidence level equals the percentage of accurate answers for that group (e.g., 90% of accurate witnesses in the group with 90% confidence). The calibration levels can be

expressed visually plotting accuracy rates as a function of confidence grouping, in this case a diagonal line would represent perfect calibration on all confidence groups (see Figure 1).

Usually, a confidence index (CI) is computed, comprising the average squared discrepancy for each witness in a given confidence level and actual proportion of witnesses who were correct in the same confidence group. Value of CI near 0 indicate a stronger confidence-accuracy relation. Another related statistic is over/underconfidence (O/U), which is computed by subtracting the mean accuracy from mean confidence for the entire witness sample, with scores that range from -1 (underconfidence) to +1 (overconfidence).

Research adopting calibration techniques show that good calibration might exist for eyewitness confidence-accuracy in identification tasks even if the data yield low correlations (r) values. Olsson (2000), for example, investigated the r , CI, and O/U values of seven experiments and found that correlations ranged from .35 to .49 ($M = .42$), while the CI and O/U values ranged from .05 and .08 and .02 to .04, respectively (values near 0 indicate better calibration). Therefore, although the correlations were small to moderate, the calibration indices were found to be more valid indicators of accuracy. Also, Brewer et al. (2002) conducted a study using a videotaped theft and found good calibration indices for eyewitness identifications (CI = .05 and O/U = -.058), even though the r for this same experiment was only .14. These and other findings repeatedly show that the CA relationship is stronger when expressed in terms of calibration computations, than when expressed using conventional r indices (Brewer & Wells, 2006; Nils Olsson, 2000; Sauer, Brewer, Zweck, & Weber, 2010; Weber & Brewer, 2004). The practical implication of this discussion is that, at first sight, calibration measures may have more substantial forensic application, because they are often better predictors of identification accuracy.

Leippe and Eisenstadt (2014) identify five reasons to be cautious when evaluating the usefulness of calibration methods, even though this approach seems to show a higher level of

a CA relationship correspondence than r . The first is that the small literature on CA calibration comprises studies that have not found good calibration in identification tasks (e.g., Cutler & Penrod, 1989; Granhag, 1997). Second, CI values present the same diagnosticity problem that modest CA r values do. A CI of .05, for example, means that a witness subjective probability of accuracy differed by a percentage of .22 (the square root of .05) on average, from the accuracy rate of their confidence group. In other words, the interpretation of a witness CI score is dependant on multiple-assessments comparisons. Of course, the comparison of multiple assessments is not likely to be present in real cases where single-assessments of witness confidence for identification of a single target predominate. The third one is that percentage scales, commonly used in experiments to assess confidence are not used by practitioners, raising doubts about comparisons between calibration and the open-ended questions adopted by police (i.e., question: “How confident are you?” answer: “Very confident”). However it is worth noting that this third limitation is of an applied practice nature, and it is not related to a problem with calibration itself. Fourth, subjective probabilities such as CI values might not be useful in actual cases, where the parties want to know for sure if a witness is correct or not. Finally, confidence and accuracy are highly malleable. This means that any type of CA relationships might still present imperfections given other social and cognitive factors that influence either confidence or accuracy.

A more complete picture of the usefulness of confidence measures can only be achieved when considering the necessary conditions for a CA relation to be stronger. In other words, CA effects are not fixed, but vary depending on many other variables that can reduce or increase the CA relationship (see Perfect, 2002; Wixted & Wells, 2017). The first studies on the role of moderator variables showed that eyewitnesses with optimal viewing conditions often show higher CA relationships (Bothwell & Deffenbacher, 1987; Deffenbacher, 1980), which was later confirmed by further studies on exposure duration (Lindsay, Read, &

Sharma, 1998; Memon et al., 2003). Lindsay, Nilsen, and Read (2000) for example, showed that CA correlations changed substantially across studies that used different encoding and testing conditions. Characteristics of the identification target such as attractiveness and distinctiveness also affect the CA relation (Brigham, 1990; Cutler & Penrod, 1989).

Eyewitnesses that take longer intervals to give a testimony are often more overconfident than those that give accounts or identifications shortly after the event, an effect possibly associated with memory contamination (Odinot & Wolters, 2006; Palmer, Brewer, Weber, & Nagesh, 2013; Sauer et al., 2010). A meta-analytic review by Sporer, Penrod, Read, and Cutler (1995) on target-present and target-absent line-ups also showed that the CA relationship was dependent on choice, with choosers (positive identifications) often having a higher CA relation than non-choosers.

Another moderator of the CA relation deserves a further scrutiny given its role in the current thesis: the differences between identification and recall tasks. Despite the extensive eyewitness literature, the vast majority of CA relation research has focused on culprit identification (e.g., line-up or photospread tasks), rather than exploring descriptions of persons and events (e.g., free recall or cued-recall tasks; Sporer, 1996). The focus on identification is understandable, considering that it might ultimately be the most conclusive information provided by a witness. However, witnesses' own freely recalled memories are also highly relevant in forensic contexts, since most of the information provided by eyewitnesses comprise descriptions of the perpetrator and the event (Van Koppen & Lochun, 1997). The small literature using eyewitness recall paradigms shows a somewhat higher CA correlation when compared to identification studies (Bornstein & Zickafoose, 1999; Odinot & Wolters, 2006; Perfect, Watson, & Wagstaff, 1993; Robinson & Johnson, 1996). In fact, recognition questions give rise to a lower CA relationship when compared to free recall memory questions that do not provide answer alternatives (Robinson et al., 1997; Robinson,

Johnson, & Robertson, 2000; Robinson & Johnson, 1996). Also, Allwood, Ask, and Granhag (2005) found good calibration and very little overconfidence using an interview recall procedure, compared to a study using the same design but with a recognition task (Granhag, Jonsson, & Allwood, 2004). Such differences between recall and recognition tasks may be accounted for by the ‘free’ or ‘self-generated’ component of recall tasks, because in such procedures witnesses tend to choose which information to provide based on their confidence, consequently increasing metamemory realism (Allwood et al., 2005). In fact, models of strategic memory reporting suggests that individuals balance the demands for informativeness and accuracy in their recalls, withholding details that fall below a preset criterion of probable accuracy (Koriat & Goldsmith, 1996; Sauer & Hope, 2016).

In the current programme of work, we provide an extensive examination of the relation between metamemory and the CA relation in eyewitness identification and free recall settings. We were particularly interested in whether eyewitnesses’ expressions of confidence are influenced by individual differences related to memory self-efficacy and endorsement of memory strategies. It was posited that if confidence judgements are partly based on individual differences and intrinsic cues (Leippe & Eisenstadt, 2014), then witnesses could tend to be underconfident or overconfident depending on their metamemory traits (Olsson & Juslin, 1999). By testing these predictions, we aimed to elucidate whether metamemory assessments could be useful to distinguish overconfident from underconfident witnesses. Importantly, this line of research also contributes to our understanding of how confidence assessments are generated in forensic relevant contexts.

Outline of the Current Thesis

This introductory chapter briefly summarizes the many implications of metamemory in eyewitness testimony research. Despite the numerous potential contributions of

metamemory frameworks to eyewitness memory research, much of this overlap has been overlooked. In the next chapters we present a line of research examining the relation between metamemory and eyewitness testimony performance. Specifically, there were three key research questions emerging from this goals. First, I examined whether there is a relation between self-reported metamemory and objective memory accuracy in different eyewitness paradigms. Our second question was whether there is a relation between metamemory and expressions of confidence in eyewitness settings. Finally, the third question focused on what is the relation between metamemory and the confidence-accuracy relationship in eyewitness performance.

In the next chapter (**Chapter 2**), we present our first set of experiments examining the relation between self-reports of memory ability and eyewitness identification performance, reporting how self-ratings of memory ability and strategies endorsement relate to eyewitness identification accuracy, confidence and over/underconfidence (Experiments 1 and 2). These experiments provide initial evidence that some metamemory factors are important indicators of eyewitness identification accuracy and confidence, although only assessments of global memory self-efficacy were adopted. In the subsequent experiments comprising this programme of work we aimed to improve our metamemory assessment by developing a measure tailored specifically to eyewitness testimony settings. In **Chapter 3** we present the development and initial tests of validation of the Eyewitness Metamemory Scale (EMS), a self-report instrument developed specifically to assess face recognition and person identification ability (Experiment 3). In **Chapter 4**, we describe an experiment designed to investigate the relation between general and eyewitness-specific metamemory assessments (EMS) on identification performance for biased and unbiased lineups (Experiment 4). We were specifically interested in possible differential effects between biased and unbiased lineups because other postdictors of identification performance (e.g., eyewitness confidence,

decision time) have less diagnostic value in biased lineups. The results of Experiment 4 suggest that some eyewitness-specific metamemory factors (as measured by the EMS) are predictive of identification accuracy for both biased and unbiased lineups. In **Chapter 5**, we adopt an innovative repeated-trials eyewitness identification paradigm to examine the relation between metamemory assessments and identification performance (Experiment 5). In **Chapter 6**, the focus is shifted from identification paradigms to an eyewitness free recall paradigm, aiming to elucidate the relation between metamemory and disclosure of information (Experiment 6). We summarize the main findings of this novel line of research in the **General Discussion**, presenting the challenges and prospects facing future eyewitness metamemory research.

CHAPTER 2

Estimating eyewitness identification performance using metamemory measures and memory tests

The research reported in this Chapter has been prepared for publication in:
Saraiva, R. B., Hope, L., Horselenberg, R., Ost, J., Sauer, J., & Van Koppen, P (under
review). Estimating eyewitness identification performance using metamemory measures and
memory tests.

Abstract

Discriminating accurate from inaccurate eyewitnesses is crucial in forensic settings, given the negative consequences of misidentifications. In some cases, confidence can be a reliable indicator of performance, but eyewitnesses can also be prone to over/under-confidence (O/U). In two experiments, we examined whether different metamemory measures and memory tests provide informative estimates of eyewitness identification accuracy, confidence, and O/U. In Experiment 1 ($N = 388$), participants first completed the Multifactorial Memory Questionnaire, the Squire Subjective Memory Questionnaire, and questions on cognitive skills. Later, in a standard eyewitness identification paradigm, participants attempted to identify a perpetrator they had seen committing a crime and provided confidence ratings. Participants then took part in one of three unrelated memory tests (cued-recall; face recognition; or general knowledge). Experiment 2 ($N = 479$) aimed to replicate the findings of Experiment 1 with more diverse stimuli. We observed three consistent findings across both experiments. First, higher contentment with general memory ability is indicative of correct identifications among choosers. Second, confidence in different memory tests are associated to confidence in eyewitness identifications, although accuracy in memory tests did not relate to identification accuracy. Third, most metamemory factors were unrelated to O/U. The results suggest that some metamemory factors are related to witness's identification accuracy, but metamemory assessments have little value in improving the diagnostic value of identification confidence. Additionally, expressions of confidence in eyewitness identification seem to be stable across other similar domains (i.e., face recognition tests).

Keywords: eyewitness memory, eyewitness identification, metamemory, memory, confidence, face recognition

Discriminating accurate from inaccurate eyewitness identifications is critical in the criminal justice system. Accurate identifications are compelling evidence for prosecuting a perpetrator, but inaccurate identifications contribute to the conviction of innocent suspects. Psychological research has identified some factors that are related to the likelihood that a culprit was correctly identified, including early statements of confidence (Brewer & Wells, 2006; Palmer, Brewer, Weber, & Nagesh, 2013; Sauer, Brewer, Zweck, & Weber, 2010; Wixted & Wells (2017), decision time during the identification (Sauer, Brewer, & Wells, 2008; Sauerland & Sporer, 2009; Weber, Brewer, Wells, Semmler, & Keast, 2004), and self-reported decision process (Dunning & Stern, 1994). However, the predictive value of these factors can vary due to a number of situational variables, and distinguishing correct from incorrect identifications is an issue far from resolved. Confidence assessments can be valid indicators of identification accuracy, but in many cases eyewitnesses are prone to be overconfident (or sometimes underconfident), making confidence measures less reliable estimates of accuracy. Evidence suggests, for example, that eyewitnesses are likely to have a weaker confidence-accuracy relationship when witnessing conditions are bad (Bothwell, Deffenbacher, & Brigham, 1987; Palmer et al., 2013), when they do not choose someone from a lineup (Sporer, Penrod, Read, & Cutler, 1995), and when biased lineups or biased instructions are used (Brewer & Wells, 2006; Leippe, Eisenstadt, & Rauch, 2009; Wixted & Wells, 2017).

From a practical perspective, a major challenge when adopting confidence as an indicator of accuracy is that different factors that can inflate or deflate confidence might be present, even though some of them can be controlled (e.g., unbiased line-up instructions; Brewer & Wells, 2006). Moreover, although the confidence-accuracy relation can be strong at the aggregate level, there are still limits about its diagnostic value for individual decisions (Sauer, Palmer, & Brewer, in press). This highlights the need for additional estimates of

accuracy, as well as the investigation of measures that can be used to discriminate dispositionally overconfident or underconfident witnesses from witnesses who might be described as ‘realistic’ in terms of the relationship between their confidence and accuracy. In this research, we tested the use of metamemory self-assessments and memory tests to predict part of the variance in eyewitness identification accuracy. Additionally, we examined whether metamemory self-assessments might be used to increase the diagnostic value of confidence in identification tasks, by discriminating overconfident/underconfident and realistic eyewitness identifications. Our rationale for testing these metamemory self-assessments and objective memory tests is that the accuracy of a specific witness – and the diagnostic value of their confidence ratings – may depend partly on individual differences factors, including those related to broader metacognitive processes.

Metamemory assessments and eyewitness identification performance

Research on Individual differences in cognitive, social and personality domains are extensive, but few studies have investigated how these differences may affect eyewitness identification performance. Most of the research in this area has focused on free recall and misinformation paradigms, showing that eyewitness susceptibility to misinformation depends on many personality and cognitive factors, such as self-efficacy, fear of negative evaluation, reward dependence, attention span, working memory and memory for faces (Andersen, Carlson, Carlson, & Gronlund, 2014; Bindemann, Brown, Koyas, & Russ, 2012; Bruck & Melnyk, 2004; Liebman et al., 2002; Morgan et al., 2007; Zhu et al., 2010). With respect to eyewitness identification, Searcy, Bartlett, and Memon (2000) found that individual differences in belief in a just world, memory self-efficacy, and perseverative errors (i.e., inability to change strategies during problem-solving) were positively related to likelihood of false identifications. However, only a small number of studies have investigated how self-

reported memory capacity might be related to eyewitness identification performance, especially focusing on metamemory components (Olsson & Juslin, 1999; Searcy, Bartlett, & Memon, 2000).

Metamemory refers to the knowledge and awareness that an individual has about his or her own memory capabilities (Dunlosky & Bjork, 2008). This introspective knowledge is often used to monitor and control one's own memory performance, including for example decisions about strategies employed to encode information (Dunlosky & Hertzog, 2000), and one's confidence that a retrieved memory is accurate (Chua, Schacter, Rand-Giovannetti, & Sperling, 2006). Metamemory judgments may be based on a variety of characteristics of memory retrieval processes, such as memory vividness and completeness (Brewer, Sampaio, & Barlow, 2005), response latency (Weber & Brewer, 2006), and the quantity and intensity of retrieved information (Koriat, 1993). Metacognitive monitoring has been studied in many different contexts, such as education, language learning, problem-solving, teaching, and eyewitness recall memory, offering an integrated theoretical framework to explain how people evaluate and judge their own memories (Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Hacker, Dunlosky, & Graesser, 2009; Koriat & Goldsmith, 1996)

In trace access theory, for example, it is postulated that both confidence and accuracy are dependent on memory trace strength. Stronger memory traces are more likely to be remembered and receive higher confidence ratings (Busey, Tunnicliff, Loftus, & Loftus, 2000; Clark, 1997). In the eyewitness domain, trace access theory is consistent by Deffenbacher's (1980) findings regarding the optimality hypothesis. That is, witnesses with better encoding conditions are more capable of making correct identifications and accurate judgments about their performance (Bothwell, Deffenbacher, & Brigham, 1987; Smith, Wilford, Quigley-McBride, & Wells, 2019). However, many important aspects of memory confidence judgements are ignored in trace access theory. Information derived from a

memory trace is just one of many sources of information that influence witnesses' metamemory and beliefs. Witnesses making a self-evaluation to gauge confidence can take into account many intrinsic, heuristic and self-credibility cues, especially when the memory trace is weak (Koriat, 1997; Leippe & Eisenstadt, 2014). Intrinsic cues include self-reflections on memory processes, from which people identify information that they learned to associate with accurate or inaccurate memory. Kebbell, Wagstaff, and Covey (1996), for example, found that witnesses showed less confidence for correct answers to hard questions when compared to correct answers to easy questions, possibly because questions difficulty is used as a heuristic to judge confidence (Wheatcroft, Wagstaff, & Manarin, 2015). Furthermore, witness's recognition speed is often related to eyewitness confidence, showing greater confidence when they are able to make rapid judgements (Robinson, Johnson, & Herndon, 1997). Heuristic cues comprise one's beliefs about external factors that can help or impair memory encoding and retrieval. It may be the case that witnesses feel more confident when they put more effort into recall tasks, even when this does not lead to changes in accuracy (Shaw & Zerr, 2003). People also usually believe that their memory will be stronger with prolonged exposure duration, consequently showing higher confidence for stimuli they have been exposed to longer (Memon, Hope, & Bull, 2003). Finally, self-credibility cues derive from people's self-evaluation of their overall memory performance. Some could overestimate their ability to recall events, being overconfident on identification and recall tasks; others may underestimate their memory ability and show underconfidence.

Studies of the relation between self-reported assessments of memory capacity and eyewitness identification accuracy are scarce. Olsson and Juslin (1999), for example, investigated whether eyewitness identification performance could be diagnosed by self-assessments of facial recognition skill, general memory skill, and encoding strategy.

Witnesses who reported that they were better-than-average face recognizers displayed greater

accuracy in an identification task than those reporting poorer face recognition skills. Moreover, the confidence-accuracy relationship was stronger for participants who rated themselves to be good face recognizers. From a theoretical perspective, the findings by Olsson and Juslin suggest that people can, at least to some extent, correctly assess their overall memory abilities using cognitive self-assessments. Consequently, such assessments might be informative when attempting to diagnose the accuracy of a given identification. However, Olsson and Juslin (1999) investigated single items that were not tested for reliability and validity, and the memory self-assessment was obtained after the recognition test, which might have influenced witnesses' evaluations of their memory abilities. We aimed to further investigate this issue using valid and reliable metamemory instruments that have demonstrated appropriate psychometric properties in a number of studies (e.g., Troyer & Rich, 2002; Van Bergen, Brands, & Jelicic, 2010; Squire, Wetzel, & Slater, 1979; Van der Werf & Vos, 2011).

Memory tests and eyewitness identification performance

In addition to the lack of studies on how self-reported assessments of memory ability relates to identification accuracy, little is known about how performance in more general memory tests relate to eyewitness identification accuracy. Although self-reports might be important predictors of eyewitness accuracy, due to the relationship between metacognitive assessments and identification performance, it might be argued that performance in memory tests would be more informative of eyewitnesses' identification accuracy. Research has shown that eyewitnesses who perform better across several intelligence and working memory tests are less susceptible to misinformation (Jaschinski & Wentura, 2002; Liebman et al., 2002; Zhu et al., 2010). Furthermore, some evidence suggests that performance in standardized face recognition tests is positively related to eyewitness identification accuracy

(e.g., Bindemann et al., 2012; Morgan et al., 2007). Thus, it might also be the case that be that individual differences in face processing and memory capacity are related to eyewitness accuracy such that unrelated memory tests can be used to estimate witness identification performance in a specific case.

However, to the best of our knowledge, there has been no research on the relation between eyewitness identification performance and performance in unrelated memory tests, such as general knowledge, cued recall, or free recall tasks. On the one hand, there is some evidence that different memory systems and memory processes are discrete, as exemplified by studies that show no effect of working memory training on improvement of other cognitive skills (Melby-Lervåg, Redick, & Hulme, 2016; Simons et al., 2016). On the other hand, it might be argued that accuracy in different memory tasks should relate to eyewitness identification accuracy, due to the stability in individual differences related to cognitive factors (Jaschinski & Wentura, 2002; Zhu et al., 2010). Some findings suggest that individuals express their confidence in similar ways across different memory tests (Bornstein & Zickafoose, 1999; Buratti, Allwood, & Johansson, 2014), which is evidence that performance in a known domain may also be related to confidence in an eyewitness context (Buratti et al., 2014). It might also be expected that performance in some tests may be more predictive of eyewitness identification accuracy than others, based on models that propose that memory is comprised of relatively independent domains (Cabeza & Moscovitch, 2013; Repovs & Baddeley, 2006). Performance on face recognition tests, for example, might be more predictive of eyewitness identification accuracy than cued recall or general knowledge tests, because the former is more closely related to an identification task (e.g., Bindemann et al., 2012; Morgan et al., 2007). In the absence of conclusive research on this topic, we sought to examine the relationship between eyewitness identification performance and performance on three different types of memory task (i.e., general knowledge, cued recall and face

recognition). A further rationale for different types of tasks is that, in practice, a face recognition task may interfere with eyewitnesses' memory of the perpetrator, so using a non-face recognition task may have greater utility when estimating eyewitness identification performance.

Experiment 1

In Experiment 1, we tested the utility of metamemory measures and memory tests to estimate witnesses' accuracy, confidence, and confidence-accuracy relation in identification tasks. We hypothesized that metamemory self-assessments of memory ability and capacity would be predictive of eyewitnesses' identification accuracy (Hypothesis 1). This prediction draws on some initial evidence that self-assessments of memory ability are related to identification accuracy (Searcy et al., 2000; Searcy, Bartlett, Memon, & Swanson, 2001; Olsson & 1999). We also expected that metamemory self-assessments would be predictive of eyewitness confidence in identification tasks (Hypothesis 2). Attempts to estimate eyewitness confidence might seem unusual, but this relates to an important theoretical question. The idea that confidence measures are affected not only by the availability of memory traces, but also by different intrinsic, heuristic and self-credibility is critical to the understanding and practical use of memory confidence in forensic contexts (Leippe & Eisenstadt, 2014). There may be important individual differences that contribute to confidence ratings over and beyond the memory trace strength, adding noise to the confidence accuracy relationship in identification tasks. Additionally, as observed by Olsson and Juslin (1999), it was expected that choosers with higher rates on metamemory scores would have higher confidence-accuracy realism in identification tasks as opposed to over/underconfidence (Hypothesis 3). If this prediction is supported, it would be evidence that metamemory assessments might be informative of the likelihood that an eyewitness is being overconfident, underconfident, or realistic about their confidence in an identification decision. It was also expected that

eyewitness identification performance would be positively related to accuracy and confidence in general knowledge tests, cued recall tests, and face recognition tests (Hypothesis 4).

Finally, it was expected that identification performance would be more closely related to the face recognition test, followed by the cued recall and general knowledge test (Hypothesis 5) on the grounds that both the identification task and face recognition task are related to face stimuli, and evidence suggests that face perception and recognition is based on domain-specific mechanisms (Kanwisher, 2000).

Method

Design. A 2 (Identification Task Target Presence: target-present vs target-absent) x 3 (Memory Test: general knowledge, cued-recall, and face recognition) between-subjects design was used. The memory tests were allocated between subjects in order to prevent contamination and fatigue effects across the different memory tests. Participants were randomly assigned to one of the three memory test conditions, and to one of the lineup type conditions. The main dependant variables were accuracy and confidence in an eyewitness identification task and in the different memory tests.

Participants. A total of 407 participants completed the study, but seven participants were excluded for failing to pass at least two out of the four attention checks, and 22 other participants were excluded because technical issues were reported (e.g., stimuli not loading properly). The final sample ($N = 388$) comprised of 81% female participants and had a mean age of $M = 21.40$, ranging from 18 to 74 years ($SD = 4.08$). The required sample size was estimated using power analysis for logistic regressions using metamemory measures as predictors of identification accuracy. We assumed that the event rate under H_0 is $p_1 = 0.5$ and that the event rate under H_1 is $p_2 = 0.65$ for $X = 1$. The odds ratio is then $OR = (0.65/0.35)/(0.5/0.5) = 1.85$, and $R^2 = 0.4$. With $\alpha = .05$ and power = 0.95, the projected

sample size needed was approximately $N = 217$. Our final sample size ($N = 388$) is larger than the minimal sample size required because data collection continued during an optimal period for recruitment. Participants either received course credits or a chance to enter a prize draw to win a £100 Amazon Voucher. The study was advertised on social media, and within a university in the United Kingdom and a university in the Netherlands.

Target event and lineups. The target event was a short film (30sec long) depicting a theft (Colloff, Wade, & Strange, 2016). In that mock crime film, a white male culprit breaks into an office, explores the room, steals a laptop, and then leaves the room. Two different line-ups with six photos were created for this stimulus, using materials generated by Colloff et al. (2016). The target-present lineup consisted of five foils and the culprit, while the target-absent lineup consisted of six foils. A pool of 40 foils was used to randomly generate the lineups for each participant, and all foils resembled the culprit. The lineups were found to be fair and unbiased in a previous study (Colloff et al., 2016).

Multifactorial Metamemory Questionnaire (MMQ, Troyer & Rich, 2002). The MMQ is an instrument with 57 items comprising three distinct factors: Contentment (i.e., affect related to memory abilities, $\alpha = 0.91$), Ability (i.e., frequency of memory problems in different situations, $\alpha = 0.88$), and Strategy (i.e., use of memory strategies in everyday life, $\alpha = 0.83$). The factor Contentment consists of 18 items (e.g., ‘my memory is worse than most other people my age’) rated on a scale from 1 (*strongly agree*) to 5 (*strongly disagree*), with higher scores indicating higher memory contentment. The factor Ability has 20 items which requires respondents to indicate how often they experienced memory mistakes over the last two weeks (e.g., ‘how often do you forget an appointment?’), on a scale ranging from 1 (*all the time*) to 5 (*never*). Higher scores in Ability indicate fewer (self-reported) memory problems in daily situations. The factor Strategy has 19 items related to the use of different memory strategies (e.g., ‘how often do you create a story to link together information you

want to remember?') and respondents indicate the frequency with which each strategy was used in the previous two weeks using a scale that ranges from 1 (*never*) to 5 (*all the time*).

Higher scores in Strategy indicate a more frequent use of memory strategies.

Squire Subjective Memory Questionnaire (SSMQ; Van Bergen et al., 2010; Squire et al., 1979). The SSMQ is an instrument that assesses development of subjective memory functioning and consists of one single factor reflecting people's beliefs about their own memory functioning ($\alpha = 0.90$). The instrument includes 18 items (e.g., 'My ability to reach back in my memory and recall what happened a few minutes ago is'), rated on nine-point scales that range from -4 (*worse than ever before*) to +4 (*better than ever before*). Higher scores in SSMQ indicate a higher improvement in self-perceived memory functioning.

Facial recognition and general memory skill assessment (Olsson & Juslin, 1999). The self-rating measures developed by Olsson and Juslin (1999) include two items that assess self-reported general memory skill (e.g., 'give an estimate of your general memory ability, compared to other people's general memory ability') and two items for self-reported facial recognition skill (e.g., 'give an estimate of your ability to remember faces as compared to other people's ability to remember faces'). Participants indicate their ability in comparison to the normal population on a 11-point scale that ranges from -5 (*much worse*) to 5 (*much better*).

Cued-recall test. For the cued-recall test, participants watched a short video (2:40min) depicting a cleaning routine in a house and then completed a questionnaire containing 17 cued-recall questions about the video (e.g., what did the woman do in the TV room? See Appendix 1). Each question was followed by a confidence scale from 0% (*not at all certain*) to 100% (*totally certain*).

Face recognition test. Forty-five adult male faces with no unusual identifying features were selected from the FEI face database (Thomaz & Giraldi, 2010). Faces were standardized in size, resolution, and background colour. During the training phase, 30 faces were randomly displayed to participants on a computer screen. Each face was presented for three seconds, with a three second inter-stimulus interval. In the testing phase, participants viewed a second set of 30 faces which included 15 faces from the training phase and 15 new faces. Each face was shown individually participants were instructed to indicate whether or not the face had been seen before and give a confidence judgement using a scale that goes from 0% (*not at all certain*) to 100% (*totally certain*). Participants had unlimited time to make their decision and proceed to the next face.

General knowledge test. A pool of 38 general knowledge questions (e.g., ‘What country is known as the Land of Rising Sun?’) was generated and pilot tested. Eight questions were considered too easy or too difficult and were removed from the final pool. Participants were asked to answer each one of the 30 questions (see Appendix 2) and rate their confidence in a scale that ranges from 0% (*not at all certain*) to 100% (*totally certain*).

Procedure. The experiment took place over two test sessions. In the first session, participants completed an online survey containing the metamemory measures (i.e., MMQ, SSMQ, and self-assessment questions of cognitive skills), and a basic socio-demographic questionnaire (i.e., age, gender, education). After 24 hours, participants received an online link to the second session. In this session, participants were informed that they would watch a short video and that they would be asked some questions later. After watching the mock crime film, they took part in a filler task for five minutes. Participants then viewed a simultaneous lineup and were asked to try to identify the perpetrator. Participants were instructed that the perpetrator may or may not be in the lineup. The photos were presented simultaneously in a 3 by 2 configuration, and the position of the culprit was determined

randomly for each subject. Participants were instructed to make a single identification or choose a ‘not present’ option, rating their confidence in a scale that ranged from 0% (*not at all certain*) and then 10, 20, 30, ... to 100% (*totally certain*). After completing the identification task, participants were assigned to one of the three possible memory test conditions described previously.

Results

Logistic regressions were conducted to test whether metamemory measures are predictive of eyewitness identification accuracy. We tested a model including confidence and all the metamemory components as predictors of identification accuracy. Two out of six diagnostic tests pointed to the presence of multicollinearity in the model, but inspection of variance inflation factor, tolerance, Farrar-Glauber F-tests, and partial correlations revealed negligible multicollinearity, so we proceeded without adopting remedial measures. Confidence was positively associated to accuracy for both choosers ($OR = 1.03, p < .001$) and nonchoosers ($OR = 1.02, p = .02$). Choosers were also more likely to be accurate if they were more content with their general memory (MMQ-Contentment; $OR = 2.13, p = .03$), but this effect was not observed for nonchoosers. None of the other metamemory factors were significant predictors of identification accuracy (see Table 2.1). Next, we fitted Bayesian logistic regression models to examine the probability that the effects of the non-significant metamemory variables are null. Models were fitted using the *rstanarm* R package, adopting standard priors that are weakly informative. Inferences were based on credible intervals, which describe the likely range of the model parameters in probabilistic terms. The results revealed that the relation between eyewitness identification accuracy and the MMQ-Ability, MMQ-Strategies, SSMQ, and facial and memory skill assessment are likely to be zero (see Figure 2.1).

Table 2.1

Logistic Regressions Testing Metamemory Measures as Predictors of Eyewitness Identification Accuracy (Experiment 1)

Choosers					Nonchoosers				
Predictor	β	$SE \beta$	p	OR [95% CI]	Predictor	β	$SE \beta$	p	OR [95% CI]
Confidence	0.03	0.01	<.001***	1.03 [1.02, 1.05]	Confidence	0.02	0.01	.02*	1.02 [1.00, 1.03]
MMQ – Contentment	0.75	0.34	.03*	2.13 [1.10, 4.26]	MMQ – Contentment	0.35	0.43	.81	1.42 [0.60, 3.35]
MMQ – Ability	-0.56	0.37	.12	0.57 [0.27, 1.16]	MMQ – Ability	-0.34	0.52	.51	0.79 [0.25, 1.99]
MMQ – Strategy	0.32	0.32	.31	1.38 [0.74, 2.61]	MMQ – Strategy	-0.22	0.39	.56	0.79 [0.36, 1.73]
SSMQ	0.06	.21	.46	1.06 [0.71, 1.61]	SSMQ	0.04	0.30	.87	1.04 [0.57, 1.91]
General Memory	0.04	.11	.70	1.04 [0.82, 1.31]	General Memory	-0.21	0.17	.23	0.81 [0.57, 1.14]
Face Memory	-0.07	0.11	0.46	0.92 [0.74, 1.13]	Face Memory	0.14	0.13	.29	1.15 [0.88, 1.50]

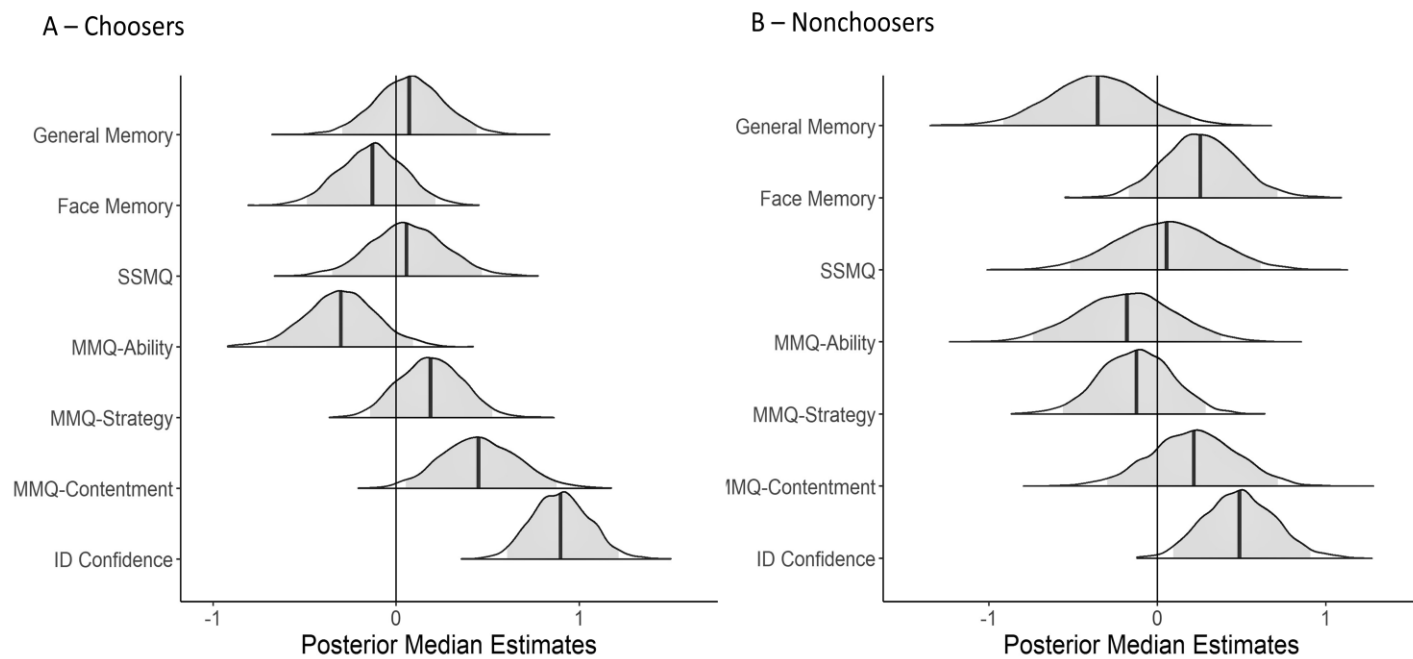


Figure 2.1. Posterior distributions of coefficients of metamemory factors as predictors of identification accuracy in Experiment 1 for Choosers (A) and Nonchoosers (B). The shaded area in the distributions represent 95% credible intervals.

We then conducted multiple regressions to test whether the metamemory measures are predictive of eyewitness identification confidence. We tested a model including accuracy and all metamemory components as predictors of identification confidence. Identification accuracy was included as a proxy for memory strength, in order to test if the metamemory measures contribute to confidence judgements beyond cues related to memory strength. As shown in Table 2.2, only identification accuracy was predictive of eyewitness identification

confidence for choosers ($\beta = 0.38, p < .001$) and nonchoosers ($\beta = 0.19, p = .02$). Results from a Bayesian regression model revealed that all metamemory factors are most probably unrelated to expressions of confidence in the identification task (see Figure 2.2).

Table 2.2

Multiple Regressions Testing Metamemory Measures as Predictors of Eyewitness Confidence (Experiment 1)

Choosers					Nonchoosers				
Predictor	<i>B</i> (SE)	β	<i>t</i>	<i>p</i>	Predictor	<i>B</i> (SE)	β	<i>t</i>	<i>p</i>
Identification Accuracy (0 = Inaccurate)	17.88 (2.83)	0.38	6.30	<.001***	Identification Accuracy (0 = Inaccurate)	9.75 (4.14)	0.19	2.35	.02*
MMQ – Contentment	-3.41 (3.21)	-0.09	-1.06	.29	MMQ – Contentment	0.53 (3.94)	0.01	0.14	.89
MMQ – Ability	2.29 (3.46)	0.06	0.66	.51	MMQ – Ability	4.47 (4.62)	0.09	0.96	0.33
MMQ – Strategy	-4.97 (2.97)	-0.11	-1.67	.09	MMQ – Strategy	3.22 (3.48)	0.08	0.92	0.35
SSMQ	0.19 (1.96)	0.01	0.10	.92	SSMQ	1.16 (2.69)	0.05	0.43	.66
General Memory	0.08 (1.09)	0.01	0.07	.94	General Memory	0.42 (1.57)	0.03	0.27	.78
Face Memory	1.54 (1.00)	0.11	1.53	.13	Face Memory	-0.03 (1.21)	-0.01	-0.02	.98

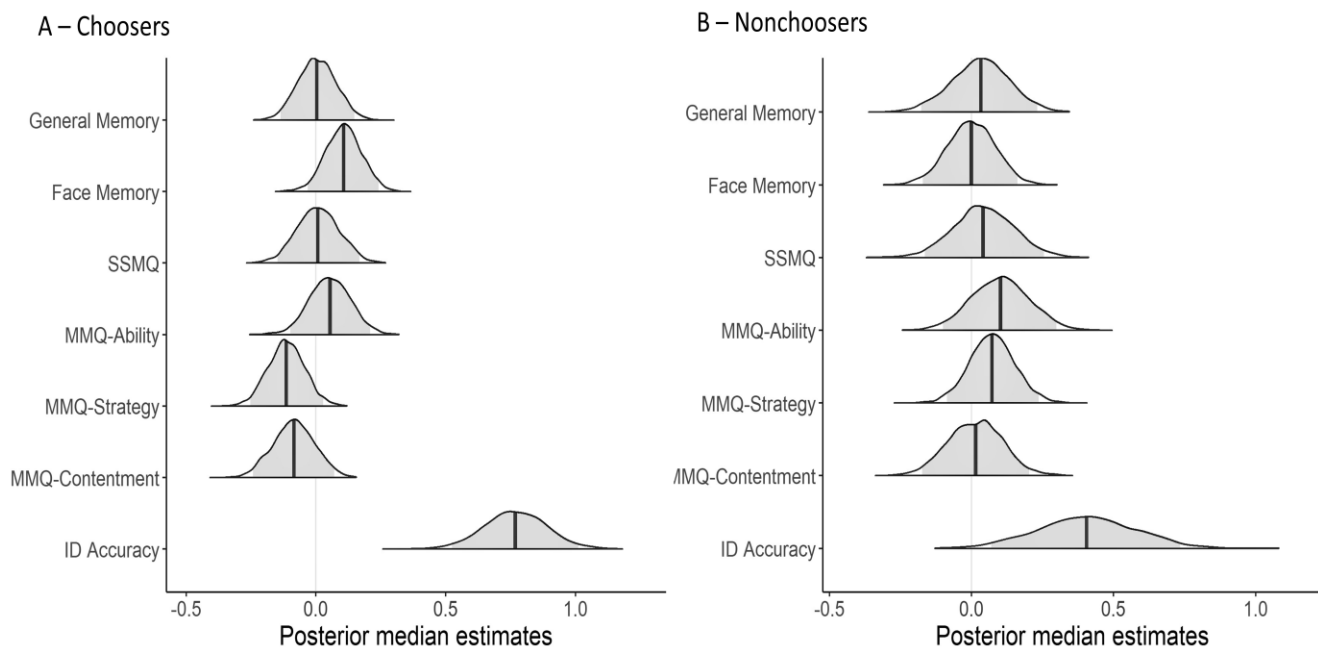


Figure 2.2 Posterior distributions of coefficients of metamemory factors as predictors of identification confidence in Experiment 1 for Choosers (A) and Nonchoosers (B). The shaded area in the distributions represent 95% credible intervals.

Calibration analyses. One of our primary aims was to investigate whether metamemory assessments were related to eyewitness confidence-accuracy calibration. In this analysis we focus specifically on choosers because it has been documented that postdictors of

identification performance have different associations for choosers vs nonchoosers (e.g., Sauerland & Sporer, 2007, 2009), and because triers of fact are more specifically concerned with eyewitnesses who choose someone from a lineup, rather than eyewitnesses who reject a lineup (Mickes, 2015). In calibration curves, the diagonal line represents perfect calibration, such that each level of confidence is equivalent to the level of accuracy for decisions made with that level of confidence (see Figure 2.3). Observations above this line indicate underconfidence, and observations below this line indicate overconfidence. We have also computed three commonly used calibration statistics: calibration index, over/underconfidence and resolution. Calibration (C) represents how far a given calibration curve is from a perfect calibration. It ranges from 0 (perfect calibration) to 1, and lower values indicate better calibration. Over/underconfidence (O/U) indicates if a curve strays more above or below the perfect calibration line, with values ranging from -1 (very underconfident) to 1 (very overconfidence). The Adjusted Normalized Resolution Index (ANRI) represents how well confidence discriminates accurate from inaccurate identifications, with higher values indicating better discrimination (see Brewer and Wells, 2006 for a detailed description and calculation of these statistics).

Following a procedure similar to the one used by Olsson and Juslin (1999), individuals above the 66th percentile were selected as high scorers and individuals below the 33th percentile were selected as low scorers. Calibration statistics and calibration curves were produced for low scorers and high scorers for each of the metamemory components. Inspection of the calibration curves showed that individuals with higher MMQ-Strategies (i.e., endorsement of memory strategies) were better calibrated than lower scorers at higher levels of confidence (see Figure 2.3). All other metamemory factors did not show any clear distinctive pattern. Following Palmer et al. (2013), we used a jackknife procedure to compute standard errors for each calibration statistic, which were then converted to 95% inferential

confidence intervals (Tryon, 2001). If the confidence intervals do not overlap, that indicates a significant difference. This analysis revealed that there are no differences in calibration between higher and lower scorers on the metamemory factors (see Figure 2.4).

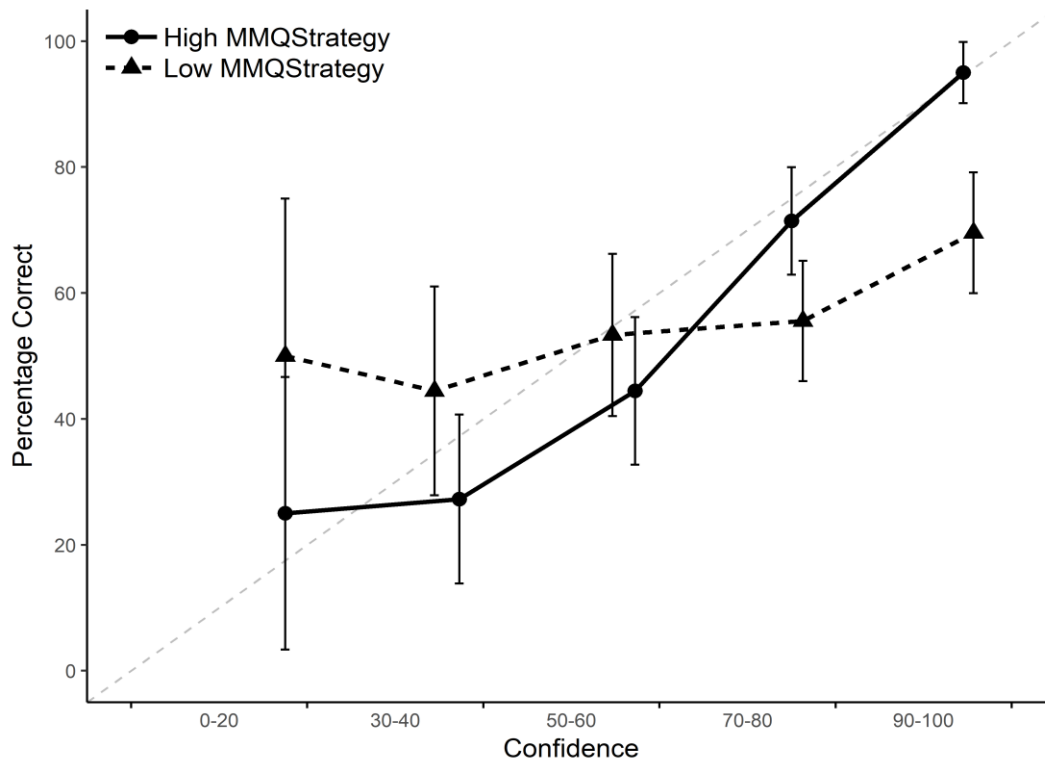


Figure 2.3. Calibration curve for Choosers identification performance by higher and low scorers in MMQ-Strategy (Experiment 1). The diagonal dotted grey line represents perfect calibration.

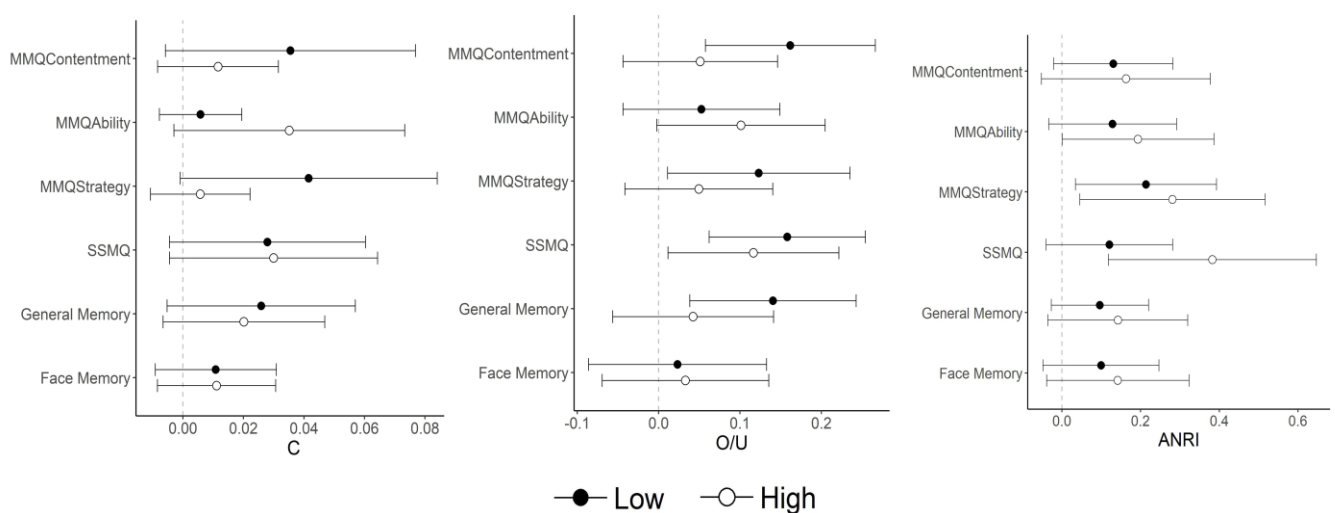


Figure 2.4. Inferential confidence intervals of calibration statistics for high and low scorers in each metamemory measure in Experiment 1.

Performance in identification task and memory tests. The next step in data analysis was to test whether performance in different objective memory tests predicted performance in an eyewitness identification task. We first tested three separate logistic regression models, with each model including accuracy in one of the memory tests (i.e., cued-recall, face recognition and general knowledge) as predictors of identification accuracy (see Table 2.3). The results showed that eyewitness identification accuracy was not predicted by performance in any of the different memory tests. Next, we used simple linear regressions to test whether confidence in the identification can be predicted by confidence in the different memory tests (see Table 2.3). The results showed that confidence in the memory tests did predict confidence in the identification task, with larger coefficients for the face recognition test ($\beta = 0.40$), followed by cued recall ($\beta = 0.25$) and general knowledge ($\beta = 0.19$).

Table 2.3.

Logistic and Multiple Regression Analyses Examining Performance in Memory Tests as Predictors of Eyewitness Identification Accuracy and Confidence (Experiment 1)

Accuracy				Confidence					
Predictor	<i>B</i> (SE)	<i>p</i>	OR [95% CI]	Predictor	<i>B</i> (SE)	β	<i>t</i>	<i>p</i>	
Model 1 – General Knowledge	0.67 (1.13)	.55	1.95 [0.21, 18.78]	Model 1 – General Knowledge	0.25 (0.11)	.19	2.27	.02	
Model 2 – Cued Recall	0.43 (1.46)	.76	1.54 [.08, 27.57]	Model 2 – Cued Recall	0.31 (0.11)	0.25	2.80	.005	
Model 3 – Face Recognition	0.86 (1.53)	.57	2.36 [0.11, 50.20]	Model 3 – Face Recognition	0.57 (0.11)	0.40	5.01	<.001	

Discussion

The results of Experiment 1 revealed four key findings. First, individuals who were more content with their own general memory were more likely to be accurate if they chose someone from a lineup than those that were less content with their general memory. Second, metamemory factors related to memory self-efficacy and endorsement of memory strategies were unrelated to expressions of confidence in identification tasks. Third, irrespective of how that confidence related to accuracy, the confidence-accuracy relationship in identification tasks was not related to intrinsic cues of memory self-efficacy and memory strategies. Fourth,

confidence expressed in the identification was somewhat stable across different memory tests, although accuracy in the identification test was unrelated to accuracy in different memory tests.

One limitation of Experiment 1 is that only one stimulus was presented during encoding, limiting the generalizability of these results to other stimuli and targets. In Experiment 2 we aimed to address this limitation by using four different mock-crime stimuli with distinct targets. This way, Experiment 2 served as a replication test for the findings obtained in Experiment 1, while also addressing issues related to stimulus sampling.

Experiment 2

Method

Participants. A total of 588 workers from Amazon Mechanical Turk completed the study online for U\$3.00. Of these, 29 participants were excluded for not passing at least two out of the four attention checks, 16 other participants were excluded because technical issues were reported (e.g., films not loading properly), and 64 other responses were excluded removed due to suspicious bot activity (i.e., Prims & Motyl, 2018). The final sample ($N = 479$) comprised of 58% male participants and had a mean age of $M = 34.01$, ranging from 19 to 70 years ($SD = 10.7$).

Materials and procedure. All experimental protocols were the same as in Experiment 1, except that in Experiment 2 four different target events were used. All target events were short films (30sec long) depicting different nonviolent crimes (see Colloff et al., 2016, for a more complete description of each crime). All identification procedures were the same as in Experiment 1, and each participant saw only one of the four different target events, and completed only one identification task.

Results and Discussion

Our analysis plan for Experiment 2 followed the same steps as in Experiment 1.

Logistic regressions revealed that Confidence was positively associated to accuracy for both choosers ($OR = 1.04, p < .001$) and nonchoosers ($OR = 1.03, p = .01$). Choosers were also more likely to be accurate if they were more content with their general memory (MMQ-Contentment; $OR = 2.37, p < .001$), but this effect was not observed for nonchoosers. None of the other metamemory factors were significant predictors of identification accuracy (see Table 2.4 and Figure 2.5). Results from multiple linear regressions revealed that identification confidence was predicted by identification accuracy, both for choosers ($\beta = 0.40, p < .001$) and nonchoosers ($\beta = 0.20, p = .009$). Higher self-perceived memory functioning (SSMQ) was indicative of higher confidence in the identification task ($\beta = 0.19, p = .002$; see Table 5). Results from a Bayesian regression model revealed that all other metamemory factors are most probably unrelated to expressions of confidence in the identification task (see Figure 2.6).

Table 2.4

Logistic Regressions Testing Metamemory Measures as Predictors of Eyewitness Identification Accuracy (Experiment 2)

Choosers					Nonchoosers				
Predictor	β	$SE \beta$	p	OR [95% CI]	Predictor	β	$SE \beta$	p	OR [95% CI]
Confidence	0.03	0.01	<.001	1.04 [1.03, 1.05]	Confidence	0.02	0.01	.01	1.03 [1.01, 1.05]
MMQ – Contentment	0.86	0.17	<.001	2.37 [1.70, 3.36]	MMQ – Contentment	0.65	0.53	.22	1.92 [0.68, 5.76]
MMQ – Ability	-0.15	0.13	.24	0.85 [0.66, 1.11]	MMQ – Ability	-0.27	0.36	.45	0.76 [0.35, 1.48]
MMQ – Strategy	-0.04	0.14	.78	0.96 [0.72, 1.28]	MMQ – Strategy	0.07	0.54	.84	1.07 [0.49, 2.39]
SSMQ	-0.03	0.08	.62	0.96 [0.82, 1.12]	SSMQ	0.01	0.21	.95	1.01 [0.66, 1.55]
General Memory	-0.07	0.07	.37	0.93 [0.79, 1.08]	General Memory	0.04	0.22	.84	1.04 [0.69, 1.63]
Face Memory	0.08	0.08	.32	1.08 [1.29, 0.33]	Face Memory	0.06	0.22	.79	1.06 [0.70, 1.71]

Table 2.5

Multiple Regressions Testing Metamemory Measures as Predictors of Eyewitness Confidence (Experiment 2)

Choosers					Nonchoosers				
Predictor	$B (SE)$	β	t	p	Predictor	$B (SE)$	β	t	p
Identification Accuracy (0 = Inaccurate)	20.83 (2.71)	0.40	7.68	<.001	Identification Accuracy (0 = Inaccurate)	13.97 (5.32)	0.20	2.62	.009
MMQ – Contentment	-0.24 (2.71)	-0.01	-1.12	.90	MMQ – Contentment	-1.89 (3.32)	-0.05	-0.57	.56
MMQ – Ability	0.67 (1.55)	0.02	0.43	.66	MMQ – Ability	0.82 (2.09)	0.04	0.39	.69
MMQ – Strategy	1.18 (1.71)	0.03	0.68	.49	MMQ – Strategy	0.22 (2.42)	0.01	0.09	.92
SSMQ	2.81 (0.89)	0.19	3.13	.002	SSMQ	0.99 (1.30)	0.07	0.76	.44
General Memory	-.58 (0.90)	-0.09	-0.64	.52	General Memory	2.08 (1.21)	0.18	1.71	.09
Face Memory	1.56 (1.01)	-0.04	-1.54	.12	Face Memory	-2.94 (1.44)	-0.21	-2.04	.04

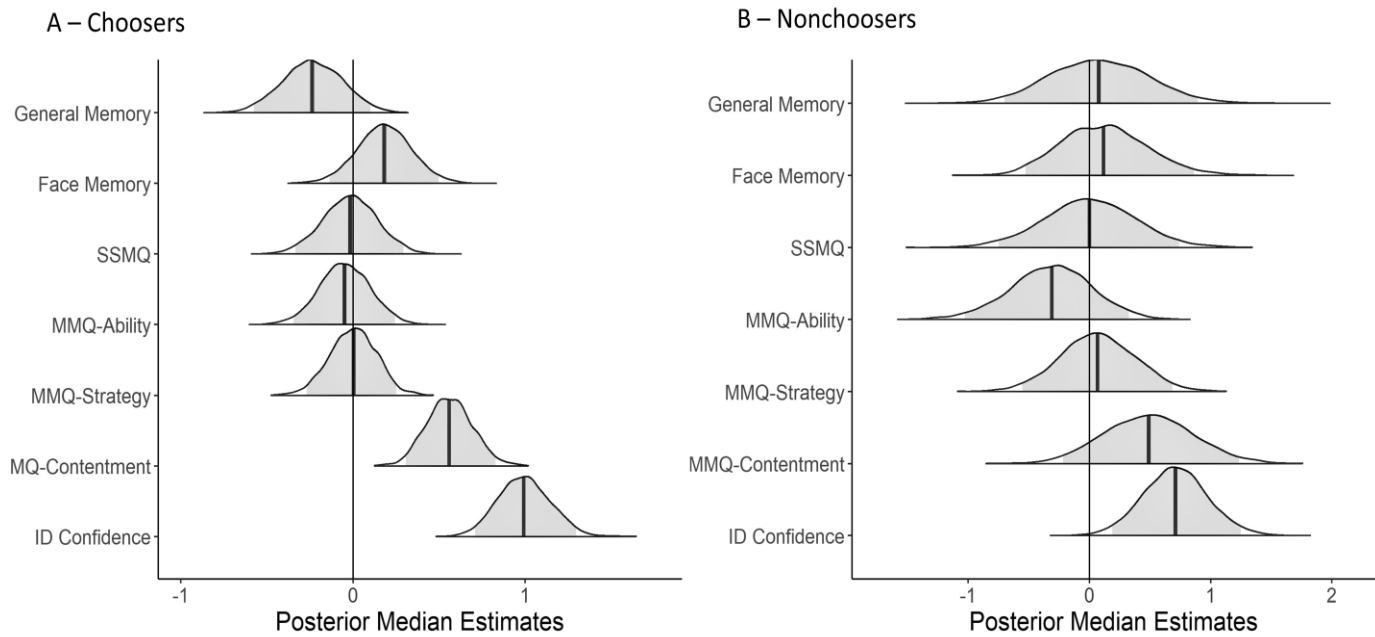


Figure 2.5. Posterior distributions of coefficients of metamemory factors as predictors of identification accuracy in Experiment 2 for Choosers (A) and Nonchoosers (B). The shaded area in the distributions represent 95% credible intervals.

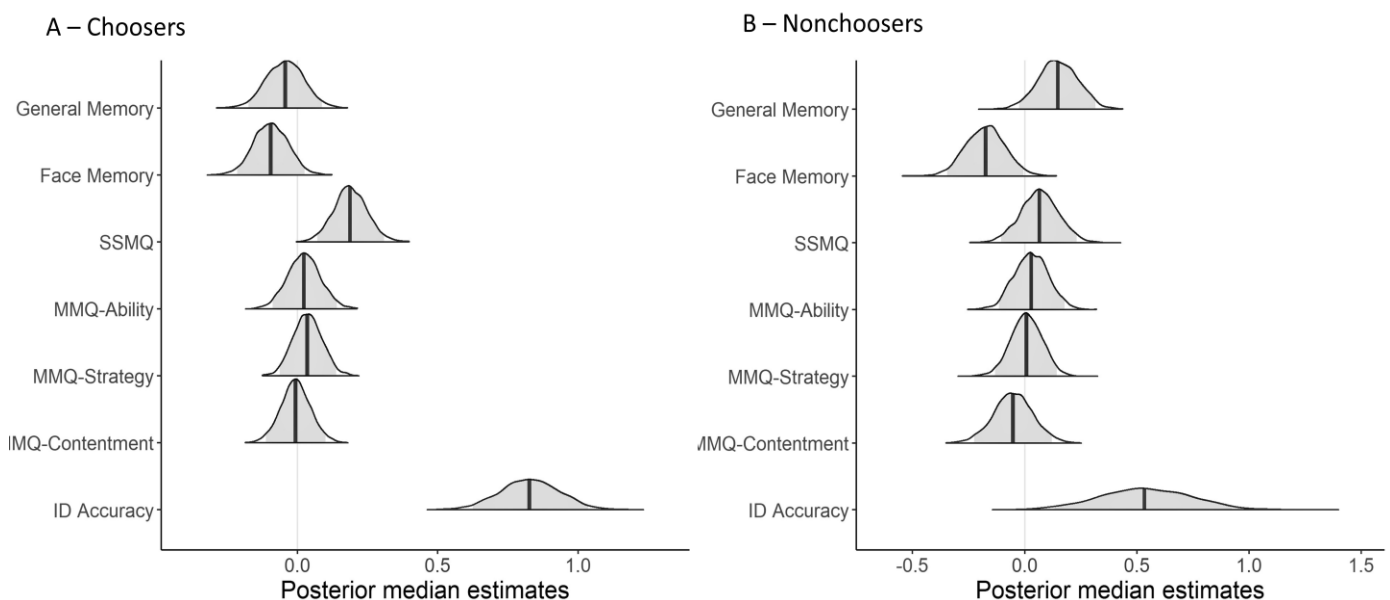


Figure 2.6. Posterior distributions of coefficients of metamemory factors as predictors of identification confidence in Experiment 2 for Choosers (A) and Nonchoosers (B). The shaded area in the distributions represent 95% credible intervals.

Calibration analyses. Inspection of the calibration curves showed that the pattern of stronger calibration for individuals with high MMQ-Strategy observed in Experiment 1 was not replicated (see Figure 2.7). Confidence-intervals computed via a jack-knife procedure confirmed these results and showed that all other metamemory factors presented no differences in their calibration (See Figure 2.8).

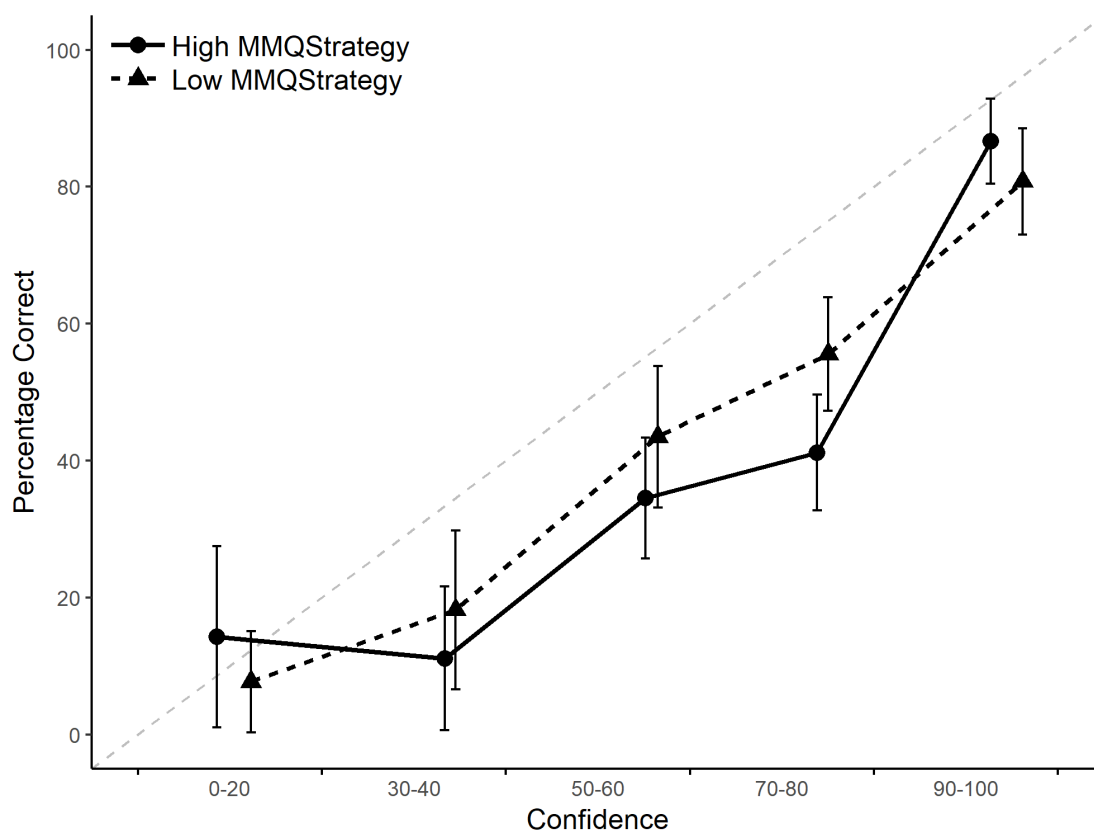


Figure 2.7. Calibration curve for Choosers identification performance by higher and low scorers in MMQ-Strategy (Experiment 2). The diagonal dotted grey line represents perfect calibration.

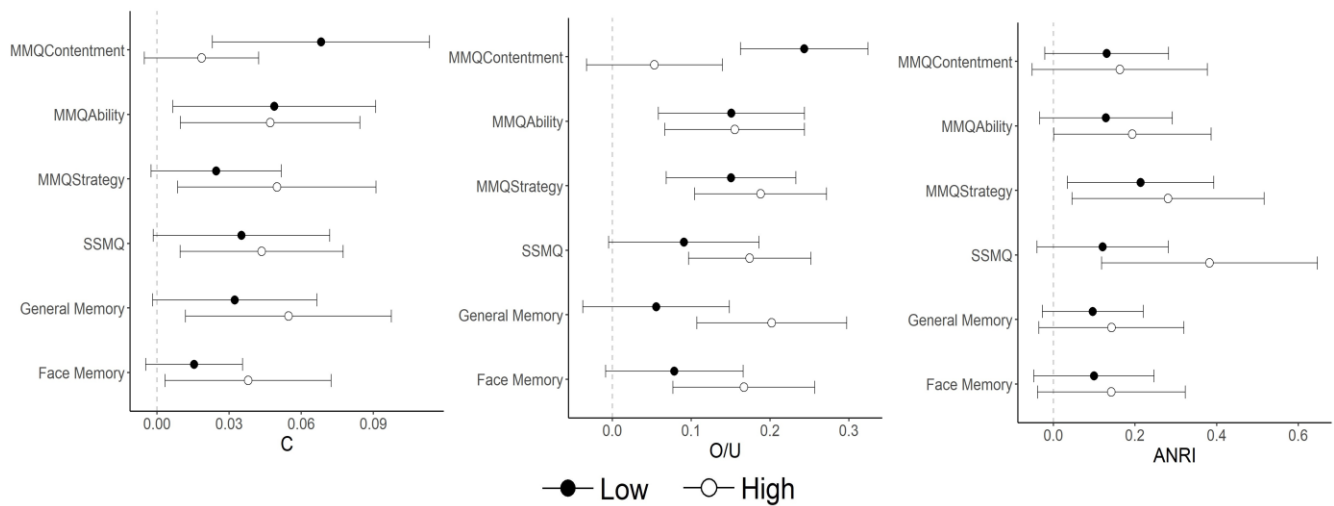


Figure 2.8. Inferential confidence intervals of calibration statistics for high and low scorers in each metamemory measure in Experiment 2.

Performance in identification task and memory tests. Results from simple regression models showed that eyewitness identification accuracy was not predicted by performance in any of the different memory tests. A separate set of regression models showed that confidence in memory tests did predict confidence in the identification task, with larger coefficients for the face recognition test ($\beta = 0.64$), followed by cued recall ($\beta = 0.30$) and general knowledge ($\beta = 0.32$).

General Discussion

In this pair of experiments, we tested the utility of metamemory measures and memory tests as estimators of eyewitness identification accuracy, confidence, and confidence-accuracy calibration. We observed three key findings that were consistent across Experiment 1 and Experiment 2. First, higher contentment with general memory capacity (MMQ-Contentment) was associated with more correct identifications among choosers.

Second, accuracy performance in different memory tests (i.e., general knowledge, cued recall and face recognition) was unrelated to eyewitness identification accuracy. Third, confidence in different memory tests was predictive of eyewitness identification confidence, with stronger relations for face recognition tests, followed by cued recall and general knowledge tests. Additionally, in Experiment 1 we find that individuals with higher endorsement of memory strategies (MMQ-Strategies) were better calibrated than individuals with less endorsement of memory strategies, although this finding was not replicated in Experiment 2.

Metamemory Assessments as Predictors of Eyewitness Identification Performance

Regarding our first main finding, we observed that contentment with one's own general memory ability (MMQ-Contentment) was predictive of identification accuracy among choosers. This result is in line with previous studies demonstrating a link between general ratings of memory self-efficacy and objective memory performance (Olsson & Juslin, 1999; Seeman, McAvay, Merrill, Albert, & Rodin, 1996; White, & McNeill, 2010). Valentijn et al. (2006), for example, found that memory self-efficacy was predictive of objective memory performance after an interval of 6 years, as measured by a visual verbal learning task. In the eyewitness identification literature, some initial evidence has been found showing a small to moderate relation between self-ratings of face recognition ability and eyewitness identification accuracy (Olsson & Juslin, 1999). In contrast, other evidence suggests that people have limited insight into their own memory ability, so that they are often unable to predict their performance in different memory tasks (Bindemann, Attard, & Johnston, 2014; Besken & Mulligan, 2013; Bobak, Mileva, & Hancock, 2018; Bobak, Pampoulov, & Bate, 2016). The current manuscript extends on this discussion by showing initial evidence that contentment with one's general memory ability (MMQ-Contentment) is related to identification accuracy among choosers. Importantly, we observe that MMQ-Contentment

and identification confidence are independent predictors of identification accuracy. That is, adopting memory self-efficacy assessments may improve estimation of identification accuracy compared to estimations that only consider identification confidence. This additional diagnostic value of memory self-efficacy might be attributed to dispositional aspects of memory performance that are not present in identification confidence. That is, confidence judgements expressed by witnesses are greatly influenced by memory trace strength (Bothwell et al., 1987; Smith, et al., 2019), while memory self-efficacy encompasses more distal experiences with previous memory performance (Hertzog & Dixon, 1994; Koriat, 1997). Therefore, assessments of memory self-efficacy may predict identification accuracy among choosers because it differentiates individuals with a more or less successful history of memory performance.

It is important to note that only MMQ-Contentment was a significant predictor of identification accuracy among choosers, while no relationship was observed for other metamemory factors related to endorsement of memory strategies, self-reported general memory and face memory ability, and memory development (SSMQ). One possible explanation that might account for our findings is that the metamemory instruments adopted are not closely related to eyewitness situations. Most of the measures used in this study focus on aspects related to general memory ability or impairment (e.g., ‘I worry that others will notice that my memory is not very good’), and in eyewitness contexts other aspects such as memory for faces are more relevant. Olsson and Juslin (1999), for example, found no evidence that general memory self-ratings were related to identification accuracy, while face memory self-ratings were somewhat related to overall accuracy. However, our analyses using the same measures adopted by Olsson and Juslin (1999) failed to replicate this finding, so that neither general nor face memory assessments were predictive of identification accuracy. As Olsson and Juslin (1999) observed, the measures they used were a combination of single

items of unknown validity and reliability. Further research on the relation between metamemory assessments and eyewitness accuracy might benefit from developing, validating and testing eyewitness-specific metamemory instruments.

A second explanation for the absence of a relationship between some metamemory assessments and identification accuracy is that people may have a limited ability to correctly evaluate their own memory performance. Perfect (2004), for example, noted that self-ratings in eyewitness memory ability were not predictive of eyewitness memory in a cued recall test. Other studies have shown that general metamnemonic self-assessments are often unrelated to actual memory ability, including difficulties in predicting the recognition of unfamiliar faces (Bindemann et al., 2014; Bobak et al., 2018). Moreover, metacognitive judgements are malleable, and can change in the presence of social influence (Semmler & Brewer, 2006). Bang et al. (2017), for example, have found that during group discussions, individuals tend to express confidence levels that match the confidence expressed by others in the group. The malleability of metamemory judgements and the dissociation between self-assessments and actual accuracy might help explain why eyewitness identification errors occur in experiments and real-life incidents. In such contexts, eyewitness can avoid identification errors by rejecting lineups or making ‘don’t know’ responses, but false identifications are common even when those options are available (Weber & Perfect, 2012). Failures to correctly assess and control for an individual’s own memory processes may well be a central underlying mechanism behind eyewitness errors, and a possible explanation of why eyewitnesses can sometimes make very confident and yet inaccurate identifications.

Another important aim of this paper was to test whether metamemory assessments might be used to estimate eyewitness identification confidence. That question was motivated by the assumption that witnesses providing confidence judgements take into account not only the strength of their memories, but also many intrinsic, heuristic and self-credibility cues

(Koriat, 1997; Leippe & Eisenstadt, 2014). We found that higher self-perceived memory functioning (SSMQ) was indicative of higher confidence in Experiment 2, but this relation was not observed in Experiment 1. Importantly, none of the other metamemory measures related to memory contentment, endorsement of memory strategies, self-reported general memory and face memory ability were significant predictors of identification confidence. Taken together, these findings seem to suggest that confidence judgements in identification tasks are not influenced by intrinsic metamemory cues related to memory self-efficacy. Therefore, we find no support for the hypothesis that individuals' insights about their own memory performance serve as a heuristic by which individuals might anchor their confidence judgements.

Some results on the link between metamemory and eyewitness confidence-accuracy relationship were inconsistent across our pair of experiments. In Experiment 1, we found that choosers with higher MMQ-Strategy scores were better calibrated compared to choosers with lower scores in this component, but this result was not replicated in Experiment 2. In Experiment 2, we recruited a larger and more diverse sample, while also employing more diverse target stimuli. This seems to indicate that the distinct calibration pattern found for the MMQ-Strategy component in Experiment 1 may have emerged due to methodological artefacts in this experiment. Overall, across both experiments we find that eyewitness confidence-accuracy calibration did not differ substantially between high and low scorers in each of the metamemory components examined. In other words, assessments of memory self-efficacy seem to have little value in improving the diagnostic value of confidence as a postdictor of eyewitness identification accuracy. Although confidence can be inflated by a number of external factors (e.g., biased lineup instructions, Brewer & Wells, 2006), our findings indicate that over/underconfidence may not be attributed to self-ratings of general memory ability.

Objective Memory Tests as Predictors of Identification Performance

Finally, we tested whether eyewitness identification accuracy and confidence could be predicted by performance in different objective memory tests, expecting stronger effects for tests that were more similar to eyewitness situations. We have found no evidence that performance on general knowledge, cued recall, and face recognition tests are good predictors of accuracy in an identification task. This finding is consistent with contemporary memory models, which often considers that memory is comprised of relatively independent domains (Repovs & Baddeley, 2006). The absence of a relationship between performance in the face recognition test and eyewitness identification is more surprising. Previous studies have found that performance in standardized face recognition tests are positively related to eyewitness identification accuracy (Bindemann et al., 2012; Morgan et al., 2007), but we did not replicate that observation in the current study. It is conceivable that other types of test reveal stronger associations between face recognition ability and eyewitness identification performance. Bindemann et al. (2012), for example, used a test more analogue to lineup identifications, comprising the selection of one single previously studied face among a number of options, instead of testing the recognition of multiple faces. Notwithstanding, we found that confidence presented in general knowledge, cued recall, and face recognition tests are significant predictors of eyewitness confidence in an identification task, with a larger effect for face recognition tests. This finding supports the assumption that confidence reports depend not only on memory vividness and trace strength, but also on internal heuristics that can be stable across different contexts. These intrinsic cues include beliefs about external factors that can help or impair memory encoding and retrieval, as well as self-reflections on memory processes, from which people identify information that they learned to associate with accurate or inaccurate memory (Koriat, 1997; Leippe & Eisenstadt, 2014).

There are a number of limitations associated with the current research. For example, the self-reports were obtained 24 hours prior to the identification task. In determining the sequence of task administration, we reasoned that exposure to the identification test before the completion of the metamemory measures might have affected witnesses' self-assessments of their memory abilities to a greater degree than completing the measures would affect eyewitness performance. While this was perhaps appropriate for an initial test of relationships between the tasks, future studies should examine the impact of collecting metamemory measures after the identification task as, in real case scenarios, such assessments would most likely be obtained after an identification procedure. We also tested metamemory measures that focus on global aspects of memory processes and experiences which are quite removed from the memorial demands placed on witnesses in forensic situations. The measures used focused mostly on aspects related to general memory ability or impairment, and more specific assessments concerning memory for faces were not thoroughly explored. Future studies might benefit from developing and testing metamemory measures that are specifically tailored to eyewitness contexts. Finally, we tested a lineup identification procedure that follows most of the recommendations for reducing eyewitnesses biased decisions. It has been argued that the confidence-accuracy relation can be weaker under biased lineup procedures (Wixted & Wells, 2017), but some studies have found only small differences in the confidence-accuracy relation for biased and unbiased identification conditions (Brewer & Wells, 2006). Therefore, it remains to be determined whether the current findings would replicate or differ when assessing witness performance on biased lineups.

Conclusion

Eyewitness evidence often provides the primary leads in criminal investigations and is of great importance during trials, especially when other evidence against a defendant is

limited or absent. However, errors committed by eyewitnesses can have severe consequences in the justice system, and discriminating between accurate from inaccurate eyewitness identifications is still one of the main challenges in this field. The confidence-accuracy relationship can be weakened due to various system and estimator factors (Bradfield, Wells, & Olson, 2002; Leippe & Eisenstadt, 2014). Individual differences measures might help discriminate individuals who are likely to have better or worse CA calibration in cases where such factors can no longer be controlled, such as after an identification was made. We found that metamemory assessments focused on contentment with general memory ability can be used to predict eyewitness identification accuracy. However, we found no evidence that metamemory assessments can improve the diagnostic value of identification confidence, one predictor commonly used in the criminal justice system. We also found that accuracy in different objective memory tests does not relate to identification of perpetrators, but confidence on such tests predicts confidence in the identification test. This finding contributes to the development of theoretical frameworks for metacognition that aim to make predictions sensitive to factors related to trace strength and internal heuristics. In conclusion, we found initial evidence that metamemory measures may provide useful information that could be considered in forensic practice, although further research is required to corroborate the reliability and generalizability of these findings.

CHAPTER 3

Development and validation of the Eyewitness Metamemory Scale

The research reported in this Chapter has accepted for publication in:

Saraiva, R. B., Van Boeijen, I. M., Hope, L., Horselenberg, R., Sauerland, M., & Van Koppen, P (in press). Development and validation of the Eyewitness Metamemory Scale. *Applied Cognitive Psychology*.

Abstract

Metamemory can be defined as the knowledge about one's memory capabilities and about strategies that can aid memory. In this paper, we describe the development and validation of the Eyewitness Metamemory Scale (EMS), tailored specifically for use in face memory and eyewitness identification settings. Participants ($N = 800$) completed the EMS and other measures on general metamemory. Results from exploratory and confirmatory factor analysis revealed good factorial validity, internal consistency, and content validity. The EMS items emerged into three distinct factors: memory contentment, memory discontentment and memory strategies. The EMS is a brief and easily administrable questionnaire that might be used to assess self-ratings of face recognition capacity and use of strategies to encode faces.

Keywords: Eyewitness Identification, Metamemory, Face Recognition, Scale Development, Factor Analysis

Metamemory can be defined as the knowledge about one's memory capabilities and strategies that can aid memory (Shimamura, 2008). This construct has been the subject of a substantial amount of research, sparked by developmental studies investigating how the ability to evaluate one's memory processes and mnemonic strategies improved learning during early childhood (Cavanaugh & Perlmutter, 1982). Metamemory research has since expanded to a variety of domains such as cognitive neuropsychology, educational psychology and cognitive psychology, motivating the development of diverse self-report measures on memory monitoring and control (Pannu & Kaszniak, 2005). However, current psychometric instruments for assessment of metamemory typically focus on broad memory domains (e.g., episodic memory or semantic memory), and there appears to be an absence of self-assessment instruments of memory capacity for faces and person recognition. In this paper, we present the development and initial validation evidence for a metamemory assessment scale tailored specifically to face memory and eyewitness identification settings.

Metamemory research is essential for a comprehensive understanding of how people use and perceive their own memory, providing a theoretical framework that can generate testable hypotheses. For example, in research examining feeling-of-knowing judgements, participants decide whether they have studied some new information sufficiently for future recall. If the subjective memory confidence experienced indicates they have not sufficiently learned the material, they may employ mnemonic strategies or engage in further study to better learn the material (Koriat, 1993). Other important branches of metamemory research include investigations on the relationship between metacognitive judgment and memory performance (Kelemen, 2000), use of memory strategies (Guerrero Sastoque et al., 2019), regulation of retrieval (Goldsmith, Pansky, & Koriat, 2014), and how metamemory changes across the lifespan (Ghetti, Lyons, Lazzarin, & Cornoldi, 2008).

Interest in assessing different aspects of metamemory has stimulated the development of various self-report measures that differ in content and item format. The content may include different aspects of metamemory so that respondents are asked to indicate the frequency of forgetting, the vividness of remembering, contentment with one's memory and perceived changes or decay in their capabilities. The item format can also vary so that some instruments focus on the relative frequency of memory issues in relation to others or in relation to one's own performance across a specified period. For example, the Metamemory in Adulthood Questionnaire (Dixon, Hultsch, & Hertzog, 1988) assesses individual's knowledge of general memory processes and tasks, frequency of memory strategy use, self-rated memory ability, perceptions of memory stability over time, anxiety regarding memory, memory and achievement motivation, and locus of control in memory abilities. The Multifactorial Memory Questionnaire (MMQ; Troyer & Rich, 2002) was developed to assess separate dimensions of memory ratings that are applicable to clinical assessment and intervention. This instrument includes scales of contentment regarding one's memory, self-appraisal of one's memory capabilities, and reported frequency of memory strategy use. Another example is the Squire Subjective Memory Questionnaire (SSMQ; Van Bergen, Brands, & Jellic, 2010; Squire, Wetzel, & Slater, 1979), assessing how one's memory trust has developed over time.

Despite the existence of several self-report memory questionnaires, there seems to be an absence of instruments that focus specifically on self-rated memory capacity for faces and person recognition. Most of the current measures have a strong focus on clinical assessments or interventions, and typically include items concerning self-evaluation of general memory ability or items concerning semantic or episodic memory issues. One notable exception is the newly developed Stirling Face Recognition Scale (SFRS; Bobak, Mileva, & Hancock, 2018). The SFRS was developed to assess face recognition ability, ranging from developmental

prosopagnosia (i.e., a neurological disorder characterized by the inability to recognize faces) to super-recognition. It has two components, face processing and face memory, which correlated moderately with objective face matching tests (correlations between $r = .28$ and $r = .34$). However, this instrument has not yet been subjected to factor analysis and the reliability of each SFRS component is unknown. Furthermore, The SFRS does not include items related to other person identification elements that may be relevant in eyewitness settings.

Self-report instruments specifically developed to measure face recognition ability and person identification would have important implications for research and practice. One important issue in the criminal justice system, for instance, is to distinguish accurate from inaccurate eyewitness identifications. Evidence obtained from witnesses of crimes can be very influential in court decisions, but inaccurate witness identifications can impair investigations and in more severe cases contribute to miscarriages of justice. Some postdictors of eyewitness identification accuracy have been identified, such as early statements of confidence (Brewer & Wells, 2006), decision time (Sporer, 1993) and decision process (i.e., absolute vs relative judgements, Dunning & Stern, 1994). However, under certain circumstances the predictive value of those factors is undermined, for example when eyewitnesses are exposed to biased lineups (Charman, Wells, & Joy, 2011) or receive feedback after an identification is made (Semmler, Brewer, & Wells, 2004). This limitation highlights the importance of investigating new factors that may be used to estimate eyewitness accuracy that are less undermined by external factors. One such potential estimator is self-efficacy in face recognition, which has shown to be predictive of eyewitness accuracy performance (Olsson & Juslin, 1999; Perfect, 2004). However, previous studies on this issue have used single items of unknown reliability and validity, limiting conclusions regarding the relation between self-efficacy and objective memory performance. A reliable and valid metamemory scale tailored specifically to eyewitness settings would improve the

inferences in studies investigating the relation between self-ratings of memory ability and objective memory accuracy.

Another important theoretical implication of an eyewitness metamemory scale is that it would help elucidate the relation between self-ratings of memory ability and expressions of confidence. Koriat (1993) has proposed that expressions of memory confidence are partly based on the encoding experience (i.e., characteristics of the stimuli) and on internal cues or beliefs about memory capacity (i.e., “am I good at recognizing this type of stimuli?”). However, general theories of memory confidence have not yet been thoroughly examined in eyewitness contexts. In forensic settings, for example, eyewitness confidence judgments are commonly used for assessing the likelihood that the eyewitness memory is accurate (Wixted & Wells, 2017). The ability to accurately evaluate one’s own memory performance is a critical feature of metamemory function, but laboratory manipulations have shown that eyewitness confidence can be inflated by factors such as post-identification feedback (Douglass & Steblay, 2006) and repeated recall (Odinot & Wolters, 2006). It has been suggested that confidence expressed by witnesses is also influenced by internal cues (Leippe & Eisenstadt, 2014), but the extent to which memory accuracy and confidence for faces is related to self-perceived recognition skill is relatively unknown. In one of the few studies on the matter, Olsson and Juslin (1999) found that people who claim to be good face recognizers show slightly higher accuracy and better confidence-accuracy calibration in eyewitness identifications, but that study is limited by the use of single items of unknown reliability and validity. The absence of valid measures of eyewitness face recognition ability impairs the advancement of this theoretical line of research. With such a measure, it would be possible to better examine the relation between beliefs of memory capacity and expressions of confidence in forensic relevant contexts.

Despite the benefits of self-report tools, it can be argued that memory accuracy could be better estimated by objective tests of memory performance. In fact, it has been proposed that tests of face recognition performance are informative estimators of proclivity to choose and identification accuracy (Baldassari, Kantner, & Lindsay, 2019; Russ, Sauerland, Lee, & Bindemann, 2018). However, in practical terms, objective tests of face recognition are more difficult to implement in applied and research settings. That is because commonly used tests of face recognition or face match ability are computerized and include many repeated trials (e.g., Dowsett & Burton, 2015; Russell, Duchaine, & Nakayama, 2009). Ideally, both objective memory tests and self-ratings of memory performance could be deployed to estimate eyewitness identification accuracy, but such approach may not always be possible due to time and resources constraints. In this scenario, brief self-ratings of memory ability may be a feasible alternative to provide estimates of accuracy in practical settings and in empirical studies, although the relation between eyewitness self-ratings of memory capacity and objective performance has yet to be elucidated (Olsson & Juslin, 1999).

In sum, a metamemory instrument tailored specifically to eyewitness settings would be of considerable value in several lines of research and has the potential to aid end-users in forensic contexts. Obtaining valid measures of metamemory for eyewitness identification is essential in research investigating the relation between self-efficacy, objective accuracy, and expressions of confidence. Depending on the results and development of this line of research, self-ratings of memory ability may also be employed to distinguish accurate from inaccurate identifications or to identify individuals with superior face recognition abilities (Russell et al., 2009). In this article, we present the development steps and initial evidence of the psychometric validity of the Eyewitness Metamemory Scale (EMS), a self-report memory instrument tailored specifically to face recognition and eyewitness identification settings. For the purposes of this study, we aimed to develop the instrument and test its factorial structure,

while also testing for its convergent and discriminant validity through associations with other metamemory measures.

Materials and Methods

Participants and Procedure

A total of 1347 participants proceeded past the informed consent page, although 143 cases were removed for failure to complete the metamemory measures. Several exclusion criteria were adopted to ensure the quality of the data: (a) 38 cases were removed for taking more than 90 minutes to complete the experiment (without outliers the study took in average 30min to be completed); (b) 145 cases were removed for completing the experiment in under 15 minutes (i.e., an impossible time to attentively complete the study); (c) 78 cases were removed for not passing all of the attention checks; and (d) 137 cases were removed due to suspicious bot activity (i.e., Prims & Motyl, 2018). The final sample ($N = 800$) comprised 62% female participants and had a mean age of $M = 29.83$, ranging from 18 to 72 years ($SD = 11.89$). The sample was from Amazon Mechanical Turk (48%), university students attending UK and Dutch institutions (32%) and participants recruited through social media (20%). Participants recruited via Amazon Mechanical Turk received U\$0.50, students received course credits, and participants recruited via social media were entered a prize draw for the prize of two £50 Amazon vouchers.

In an online survey presented via Qualtrics, participants first completed the EMS, followed by other general metamemory scales. The EMS scale was always shown first, while the other metamemory scales were presented in a random order.¹ Demographic information

¹ Participants then took part in an eyewitness paradigm consisting of a mock crime video and two identification tasks with confidence judgements. This data was obtained as part of a larger research project aiming to investigate the relation between metamemory measures and eyewitness memory performance. Due to space and focus, we only report on those measures that are relevant to the development of the eyewitness metamemory scale.

including gender, age and level of education was also obtained and on completion of all tasks, participants were debriefed and thanked for their participation.

Materials and Instruments

Eyewitness Metamemory Scale. Two qualitative approaches were adopted to develop an initial pool of items for the Eyewitness Metamemory Scale (EMS). First, we closely examined the items of other metamemory measures and, where possible, based our item development on these items. Then a semi-structured interview was conducted with a group of legal psychologists and graduate students working in this field of research ($N = 14$) to obtain additional information regarding memory self-assessment in eyewitness contexts. The initial pool of items consisted of 35 items, including eyewitness specific items and items concerning facial recognition adapted from various metamemory questionnaires. All items were rated on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). We did not establish specific hypotheses concerning the factorial structure that would emerge from these items, but rather used an exploratory approach to establish its factorial structure.

General metamemory instruments. In addition to the EMS, participants also completed the Multifactorial Memory Questionnaire (MMQ; Troyer & Rich, 2002), and the Squire Subjective Memory Questionnaire (SSMQ; Van Bergen et al., 2010; Squire et al., 1979). The MMQ has three sub-scales: Contentment, Ability, and Strategy. All items are measured on a 5-point Likert scale. The contentment scale has 18 items (e.g., “I am generally pleased with my memory ability”; $\alpha = 0.92$) rated from 1 (*strongly disagree*) to 5 (*strongly agree*), with higher scores indicating higher memory contentment. The ability scale has 20 items related to experiences with common memory errors over the past two weeks (e.g., “how often do you forget an appointment?”; $\alpha = 0.92$) from 1 (*all the time*) to 5 (*never*), with higher scores indicating better self-reported ability. The strategy scale has 19 items concerning the use of memory strategies during the past two weeks (e.g., “how often do you

use a timer or alarm to remind you when to do something?”; $\alpha = 0.88$). The items are assessed on a scale ranging from 1 (*never*) to 5 (*all the time*), with higher scores indicating greater use of memory strategies. The MMQ has shown good high test-retest reliability and high internal consistency in the original study by Troyer and Rich (2002) and in adaptations to different countries (e.g., Fort, Adoul, Holl, Kaddour, & Gana, 2004; van der Werf & Vos, 2011). The SSMQ consists of 18 items related to Memory Trust (e.g., “my ability to recall things when I really try is”; $\alpha = 0.94$). Participants rated the items on a 9-point scale ranging from -4 (*worse than ever*) to 4 (*better than ever before*). This instrument has shown good psychometric properties in different studies and has been correlated in a meaningful way with age, cognitive failures, and susceptibility to misinformation (Van Bergen et al., 2010; Van Bergen, Horselenberg, Merckelbach, Jelicic, & Beckers, 2010). The MMQ and SSMQ differ mainly in response format. Although both instruments tap into self-rated memory ability, the MMQ focuses on present ability (i.e., “I am generally pleased with my memory ability”), while the SSMQ focuses on memory development over time (“my memory ability is better than ever before”). Those instruments were selected to test convergent and divergent validity of the EMS due to their good psychometric properties and high content validity in assessing metamemory traits such as self-ratings of memory capacity and memory trust. However, no specific hypotheses were established a priori concerning the specific relation between each of the SSMQ and MMQ factors with the factors obtained for the EMS, given that the factorial structure of the EMS was unknown prior to our analysis. Therefore, the convergent and divergent analysis in this study were exploratory, and it was generally expected that factors in the EMS would relate meaningfully with factors from the MMQ and SSMQ given the similarities between those instruments in assessing metamemory.

Attention checks. Three attention checks were included within the metamemory assessment, in which participants were asked to select a specific response for that item such

as “for this question, please select the option 4 (*better than ever before*)”. The attention checks were included as an exclusion criterion (see Participants and Procedure section).

Results

To examine the validity of the EMS factorial solutions, a within-sample replication strategy was adopted, and the total sample was randomly split in half (Osborne & Fitzpatrick, 2012). The first half was treated as a training dataset for obtaining an initial factorial solution via exploratory factor analysis (EFA). The second half was treated as a test dataset for examining the fit of the initial solutions obtained in the training dataset via confirmatory factor analysis (CFA). All analyses were performed in the statistical software package R (2019). The dataset and data analysis script can be found in Electronic Supplemental Materials 1, 2 and 3 (doi: 10.1002/acp.3588).

Exploratory Factor Analysis

Prior to the analysis, one item was removed because of a semantic error in the survey. A correlation matrix of the remaining 34 items was screened to identify items that were poorly correlated with the others, or items that were highly correlated and generating multicollinearity issues. Eight items were excluded for showing weak item-total correlations ($r < .30$). Two other items were excluded for presenting high correlations ($r > .65$) and redundant content in relation to other items.

Diagnostic tests were performed on the remaining 24 items to examine the assumptions for EFA. Data gathering for the metamemory measures was performed in an online setting with forced responses, so no missing responses were present. Graphical inspection and significant Shapiro Wilk tests for all the items indicated significant univariate non-normality, with skewness ranging from -0.87 to +0.89, and kurtosis ranging from -1.01 to +0.32. This observation was supported by the statistically significant Mardia's test, indicating that the assumption of multivariate normality was violated. Therefore, a weighted

least squares extraction method for EFA was used, which provides standard errors and tests of model fit that are robust to the non-normality of the data. The items showed good factorability (Kaiser-Meyer-Olkin test = 0.89 and significant Bartlett's test) and did not present multicollinearity or singularity issues (Determinant > 0.00001).

Parallel analysis and scree plots were used as factor retention criteria and suggested the presence of four factors. A four-factor solution was extracted using oblimin rotation to allow for correlations between the factors. This solution revealed four distinguishable factors, but one factor had only four emerging items which appeared to be related to memory development over time. These items presented high cross-loadings with two of the other factors in the solution, indicating that a four-factor solution might not be robust. We proceeded with the extraction of a three-factor solution using oblimin rotation. Examining the pattern matrix, we decided to exclude one item from the first factor for high cross-loadings and a content that was dissonant with the other items (i.e., "People are generally good at remembering unfamiliar faces"). The same three-factor extraction was then repeated on the remaining 23 items (the pattern matrix for this solution is presented in Table 3.1). Items had high loadings on their respective factor, with no cross loadings higher than .30. We termed the three factors Memory Contentment (10 items explaining 19% of the total scale variance), Memory Discontentment (8 items explaining 15% of the variance), and Memory Strategies (5 items explaining 10% of the variance). The Memory Contentment factor combined items related to positive self-perception of memory ability, including keywords such as 'satisfied', 'confident', and 'better'. The Memory Discontentment factor combined items related to negative self-perception of memory ability, including keywords such as 'trouble' and 'worse'. The Memory Strategies factor combined items related to the use of memory strategies in the context of person identification and could be defined as the extent to which an individual adopts strategies to better recognize someone in the future. Reliability of the

factors was examined using omega coefficients instead of alpha, given that assumptions for alpha are rarely met in psychometric research (Dunn, Baguley, & Brunsden, 2014). Good reliability scores were found for the Memory Contentment [$\omega = .88$, 95% CI (.86, .90)], Memory Discontentment [$\omega = .86$, 95% CI (.83, .88)], and Memory Strategies factors [$\omega = .82$, 95% CI (.78, .85)]. A separate factorial structure with two factors was also extracted for comparison purposes.

Confirmatory Factor Analysis

The purpose of the subsequent analysis was to confirm the factor structure for the 23-item EMS scale on a separate subset of our sample. The results from the EFA indicated that a three-factor solution was the most appropriate to describe the EMS. A two-factor structure was also submitted for analysis as a plausible competing model for comparing fit indices. This two-factor solution was fitted to further examine whether the Contentment and Discontentment factors in the three-factor solution emerged due to phrasing method rather than to the constructs the factors represent (Podsakoff, MacKenzie, & Podsakoff, 2012). Confirmatory factor analyses were conducted to test both models. Goodness of fit was evaluated using the robust root mean square error of approximation (RMSEA) and its 90% confidence interval, robust comparative fit index (CFI), robust Tucker-Lewis index (TLI), and Expected Cross-Validation Index (ECVI). These fit indices provide different types of information (i.e., absolute fit, fit adjusting for model parsimony, fit relative to a null model), and when combined they provide a reliable and conservative evaluation of model fit (Schreiber, Nora, Stage, Barlow, & King, 2006). The chi-square test is reported, but not relied upon to evaluate model fit due to its oversensitivity to sample size and the fact that it tests for perfect fit. The evaluation of the models was based on (i) conventional criteria for good model fit ($RMSEA < .08$, $CFI > .90$, $TLI > .90$, smallest ECVI), and (ii) the

interpretability of the solution (i.e., the comprehensibility of the factors on a conceptual level).

Table 3.1

Item-total correlations, communalities and pattern matrix for the Eyewitness Metamemory Scale

Scale items	ITC	h ²	Factor Loadings		
			1	2	3
<i>Items Relating to Memory Contentment</i>					
1. My ability to remember faces is much better than other people's ability to remember faces.	.33	.67	.84	-	-
2. I am confident with my ability to remember faces in a stressful situation.	.43	.57	.74	-	-
3. Compared to other people, I think I would be a much better eyewitness.	.40	.60	.73	-	-
4. If I saw someone commit a crime, I am certain that I would remember his/her face.	.45	.55	.70	-	-
5. If I witnessed a robbery, I would be able to recognize the perpetrator a month later.	.54	.46	.65	-	-
6. As I age, I find my ability to remember faces is getting better.	.58	.42	.65	-	-
7. I am generally satisfied with my ability to remember the faces of people I have only met once.	.50	.50	.61	-.24	-
8. My ability to correctly remember where and when I saw a particular face has improved over time.	.62	.38	.57	-	-
9. I can give a detailed description of a friend's face to a stranger.	.74	.26	.37	-.10	.20
10. I recognize relatively unknown actors, if I have seen them in another movie/television show.	.80	.20	.34	-.16	.11
<i>Items Relating to Memory Discontentment</i>					
11. Sometimes I have trouble recognizing a person that I know relatively well.	.44	.56	.26	.77	-.11
12. My ability to remember faces is much worse than other people's ability to remember faces.	.35	.65	-.19	.75	-
13. It often happens that a person who seems familiar starts a conversation with me, but I have no idea who the person is.	.54	.46	.11	.68	-
14. Whenever I meet an important person, I am worried that I will not be able to recognize him/her a week later.	.53	.47	-	.67	-
15. Compared to other people, I think I would be a much worse eyewitness.	.49	.51	-.25	.62	-
16. My ability to correctly remember where and when I saw a particular face has deteriorated over time.	.62	.38	-.10	.57	.12
17. When I see a person that looks familiar, I often do not know where I have seen that person before.	.66	.34	-.11	.55	-
18. As I age, I find my ability to remember faces is getting worse.	.65	.35	-.12	.54	-
<i>Items Relating to Memory Strategies</i>					
19. Compared to other people, I more often use a strategy (e.g., focus on specific facial features such as eyes) to remember a person's face.	.32	.68	-	-	.76
20. Compared to ten years ago, I more often use a strategy (e.g., focus on specific facial features such as eyes) to remember a person's face.	.36	.64	-	.14	.75
21. In order to remember a perpetrator's face, I would definitely use a strategy (e.g., focus on specific facial features such as eyes) to remember the perpetrator's face.	.56	.44	-	-.10	.70
22. I often focus on specific facial features such as nose and eyes when I am paying attention to a face that I have to remember.	.53	.47	-	-.10	.67
23. I often create a visual image in my mind of a face that I want to remember.	.73	.27	.18		.41
Eigenvalues			4.36	3.55	2.41
Percentage of Variance Explained			.19	.15	.10

Note. ITC = Item-Total Correlations; Factor loadings higher than >.40 are presented in bold.

Diagnostic tests were performed on the 23 items to examine the assumptions for CFA and indicated that the assumption of multivariate normality was violated. Therefore, we estimated parameters in CFA using a maximum likelihood estimation with robust standard

errors, which provides tests of model fit that are robust to the non-normality of the data (Finney & DiStefano, 2006). Figure 1 presents the model specification and goodness of fit indices for the three-factor model and two-factor model. The model fit indices suggested that the three-factor solution had a better fit compared to the two-factor solution. However, the three-factor solution did not fit the data particularly well (e.g., RMSEA > 0.08, CFI < 0.90, TLI < 0.90). In an exploratory approach, we revised the three-factor model by evaluating its modification indices, adopting only theoretically sound modifications to avoid over-specification of the model. Following this approach, we included two new correlations between errors of items 6 and 8, and items 16 and 18. These modifications were based on the content of the items, which seem to be closely related to memory development over time (e.g., “As I age, I find my ability to remember faces is getting better”). The revised model resulted in an acceptable fit to data (see Figure 3.1).

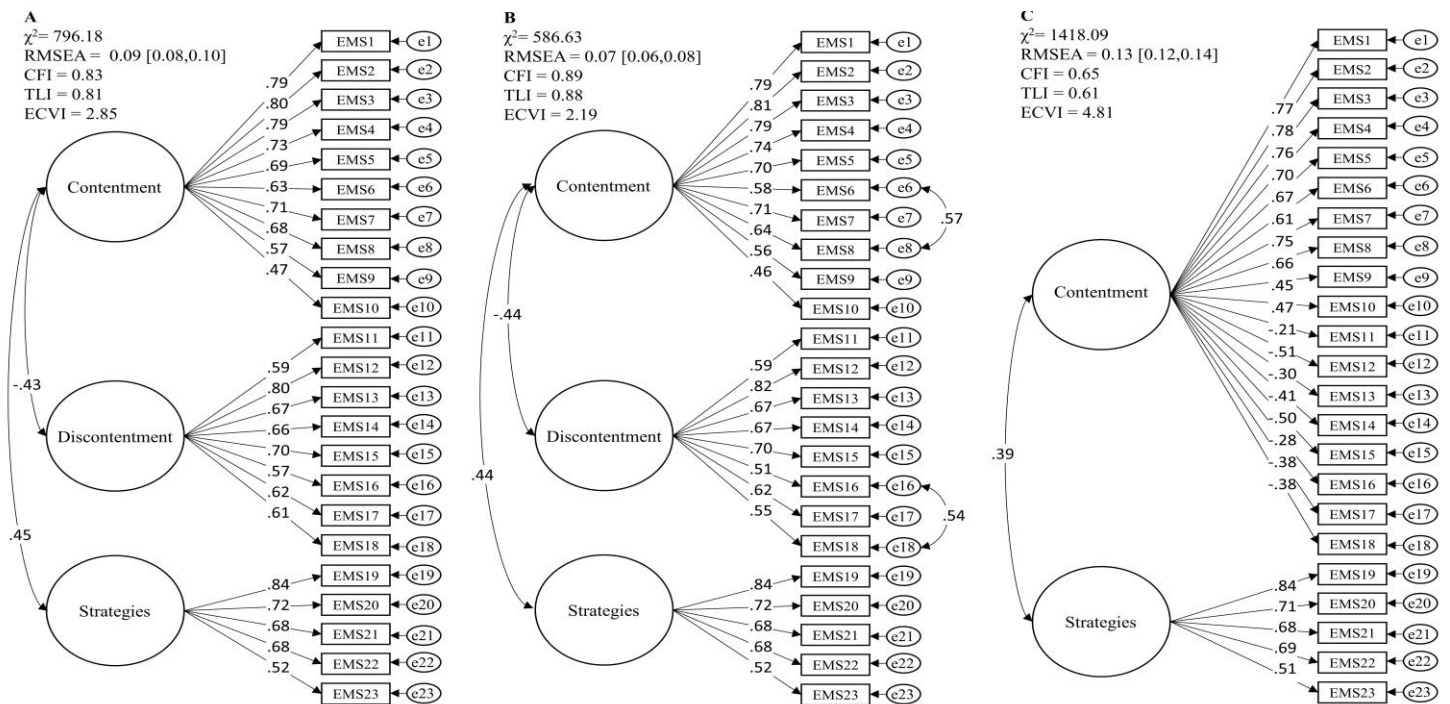


Figure 3.1. Model specification and model fit indices for the three-factor solution (A), revised three-factor solution (B) and two-factor solution (C) of the Eyewitness Metamemory Scale.

Convergent and Discriminant Validity

Convergent and discriminant validity was examined by correlating the factors of the EMS with factors from other established metamemory measures (i.e., MMQ and SSMQ). This analysis was conducted on the complete dataset ($N = 800$) instead of the training or testing dataset, given that those subsets were used only to perform independent EFA and CFA analyses. The three-factor solution demonstrated better validity in both the EFA and CFA, so the three factors emerging from this solution (i.e., EMS-Contentment, EMS-Discontentment, and EMS-Strategies) were used in the correlation analyses. Table 2 presents the correlation matrix of the EMS factors and other metamemory measures. A small negative correlation was found between EMS-Contentment and EMS-Discontentment ($r = -.29$). A moderate positive correlation was found between the EMS-Contentment and EMS-Strategies factors ($r = .44$), but no relation was found between EMS-Discontentment and EMS-

Strategies ($r = .04$). The EMS-Contentment and EMS-Discontentment factors presented a small to moderate relation with most of the other metamemory measures, while the EMS-Strategies factor presented a positive relation with MMQ-Ability, MMQ-Strategy and SSMQ-Memory Trust.

Table 3.2

Means, standard deviations, and correlations with confidence intervals of the EMS factors and other metamemory measures

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6
1. EMS-Contentment	4.22	1.02						
2. EMS-Discontentment	3.53	1.09	-.29** [-.35, -.23]					
3. EMS-Strategies	4.51	1.10	.45** [.39, .50]	.04 [-.03, .11]				
4. MMQ-Contentment	3.62	0.69	.26** [.19, .32]	-.54** [-.59, -.49]	.07 [-.00, .14]			
5. MMQ-Ability	3.57	0.62	.31** [.24, .37]	-.34** [-.40, -.27]	.19** [.12, .25]	.53** [.48, .58]		
6. MMQ-Strategy	2.86	0.64	.09** [.02, .16]	.23** [.16, .30]	.20** [.13, .26]	-.29** [-.36, -.23]	-.40** [-.46, -.34]	
7. SSMQ	5.91	1.21	.62** [.57, .66]	-.20** [-.26, -.13]	.38** [.32, .44]	.45** [.39, .50]	.42** [.36, .47]	.07* [.01, .14]

Note. Values in square brackets indicate the 95% confidence interval for each correlation. * indicates $p < .05$. **indicates $p < .01$.

Discussion

In this article, we present a new self-report metamemory questionnaire developed specifically for face recognition and eyewitness identification settings. Overall, the analyses revealed good evidence of factorial validity, internal consistency and both convergent and discriminant validity. Exploratory factor analysis yielded three meaningful factors, each of which is associated with high loadings by the items on one factor but not on the other. This

pattern of loadings provided an initial factorial validity that was also confirmed in a subset sample, corroborating the instrument division into three scales representing different dimensions of eyewitness metamemory (i.e., Memory Contentment, Memory Discontentment and Memory Strategies).

Convergent validity was demonstrated by small to large correlations between the EMS factors and other scales on multidimensional metamemory questionnaires. EMS-Contentment was positively related to self-perceived contentment and ability for general memory capacity (MMQ) and had a large correlation with memory trust (SSMQ; $r = .62$). The EMS-Discontentment factor was negatively related to self-perceived contentment and ability for general memory capacity. Interestingly, some divergent pattern of results can be observed between EMS-Contentment and EMS-Discontentment. EMS-Contentment and EMS-Discontentment were not strongly related, and a model aggregating both factors in a single memory contentment factor presented poor fit to the data in this study. Furthermore, eyewitness memory contentment was positively related to the use of strategies for person identification ($r = .45$), but this relation was not observed for eyewitness memory discontentment ($r = .04$). It may be the case that individuals with higher contentment with their own memories seek additional strategies to maintain performance, or that adopting strategies to better recognize someone result in higher satisfaction with one's memory capacity (Meinhardt, Persike, & Meinhardt, 2014). These findings indicate that, at least in part, contentment and discontentment with one's own capacity for face and person recognition may represent independent constructs, rather than opposite ends of the same spectrum. The EMS-Strategies factor had only small to moderate correlations with self-rated contentment and ability for general memory capacity and memory trust. This divergent correlation pattern seems related to the fact that EMS-Strategies items focus specifically on

the use of memory strategies to encode and remember faces, which appears to be somewhat independent from the use of strategies and self-appraisal for general memory.

Regarding discriminant validity, part of the non-shared variance between EMS and the other scales may be due to differences in content and item format. The EMS focuses specifically on memory for faces and person identification, while the other measures have a broader scope of items related to different memory domains. Contemporary memory models often consider that memory is comprised of relatively independent domains (Repovs & Baddeley, 2006), and there is some evidence that people have distinct self-perceived capacity for different memory domains (Tonković & Vranić, 2011). In terms of item format, most items in the EMS are responded in relation to present contentment with memory, while in the SSMQ, for example, items are responded in relation to memory development over time (e.g., ‘better than ever before’). The SSMQ memory development focus may be especially appropriate in clinical contexts, where changes in memory perception can indicate the advancement of medical conditions (Mitchell, 2008).

Due to space and focus, we report in the current paper the development and evidence for factorial, convergent and divergent validity of the EMS. In two other studies, it was observed that some of the EMS factors are related to eyewitness identification performance (Saraiva et al., 2019a; Saraiva, Van Boeijen, Hope, Horselenberg, & Van Koppen, 2019b). In Saraiva et al. (2019a), it was found that lineup choosers (i.e., witnesses who select someone from a lineup) with higher EMS-Discontentment were more likely to be inaccurate, in both biased ($OR = 0.57, p < .001$) and unbiased lineups ($OR = 0.56, p < .001$). In Saraiva et al (2019b), it was found that for each unit increase in EMS-Contentment score, the odds of making a correct identification increased by a factor of 1.41 ($p < .001$), and the odds of making a false identification decreased by a factor of 0.79 ($p = .002$). In both studies, it was also observed that EMS-Contentment and EMS-Discontentment were significant predictors

of identification confidence, suggesting that expressions of confidence are partially influenced by self-ratings of face recognition ability. Taken together, these studies provide initial evidence for the content validity and predictive validity of the EMS.

The EMS fills an important gap in the literature on face recognition and eyewitness testimony that might benefit a variety of research subjects. Eyewitnesses in criminal cases, for example, can provide unique evidence that may help solve investigations, hint to primary suspects, or potentially identify a perpetrator (Benton et al., 2006). However, eyewitness memory is malleable and susceptible to contamination, which may impair investigations or in more severe cases contribute to wrongful convictions. Discriminating accurate from inaccurate eyewitnesses is a challenging issue, but some predictors of eyewitness identification accuracy have been identified, such as decision time during identifications (Sauer, Brewer, & Wells, 2008), self-reported decision-making process (Sauerland & Sporer, 2007) and early statements of confidence (Wixted & Wells, 2017). Metamemory judgements and individual differences in face recognition capacity may also relate to eyewitness performance, but this hypothesis has been relatively unexplored. Some studies have suggested that people have only moderate insight into their face recognition and face perception abilities (Bobak et al., 2019), but expressions of confidence may have a stronger relation to self-perceived memory ability. This is of importance because confidence statements are often used to discriminate accurate from inaccurate witnesses, but little is known about whether confidence statements are affected by individual differences (Leippe & Eisenstadt, 2014). Research adopting self-report instruments of face recognition capacity such as the EMS could help clarify the relationship between past-experiences with memory and confidence judgements.

Research on prosopagnosia and super-recognizers could also benefit from the use of self-rated measures of face memory capacity. From a theoretical perspective, it is not clear

whether prosopagnosia and superior face recognition represent opposite ends of the same continuum of face memory ability (Bobak et al., 2019). Comparing self-reported scores of face recognition capacity on patients with prosopagnosia and super recognizers could help clarify whether objective memory capacity has a linear relation with subjective memory experience. From an applied perspective, super-recognizers are considered as particularly valuable to national security agencies and border control due to their extraordinary ability to match and recognize faces from video footage or line-ups (Bobak, Dowsett, & Bate, 2016). Valid face memory capacity self-report questionnaires could be used as screening tools among many participants prior to other behavioural testing, helping identifying individuals with remarkable face memory skill.

The EMS is a brief, easily administrable metamemory questionnaire focusing on face recognition and eyewitness contexts. It has been developed on a large and relatively heterogeneous sample and showed good psychometric properties, although future amplification of its validity is desired. Self-report assessments of memory add a unique element to the assessment of memory performance that cannot be obtained in objective memory testing alone. Self-report tools allow the measurement of overarching memory issues and experiences rather than artificial laboratory-based memory problems, providing insights on an individual's memory functioning. Consequently, such tools have an important role in research and theory development regarding how memory performance relates to one's theory and one's previous experiences with memory.

CHAPTER 4

Eyewitness metamemory predicts identification performance in biased and unbiased lineups

The research reported in this Chapter has been prepared for publication in:
Saraiva, R. B., Van Boeijen, I. M., Hope, L., Horselenberg, R., Sauerland, M., & Van Koppen, P (under review). Eyewitness metamemory predicts identification performance in biased and unbiased lineups.

Eyewitnesses play a major role in the criminal justice system, especially in cases lacking other physical evidence. In many jurisdictions, suspects are more likely to be prosecuted if an eyewitness identifies them as the perpetrator of a crime. However, as with other types of evidence, eyewitness identifications can be in error or contaminated (Wixted, Mickes, & Fisher, 2018). Researchers have identified some factors that can be used to distinguish accurate from inaccurate witnesses, including early statements of confidence (Brewer & Wells, 2006; Wixted & Wells, 2017), decision time during the identification (Sauer, Brewer, & Wells, 2008; Sauerland & Sporer, 2009; Sporer, 1993) and self-reported decision process (Dunning & Stern, 1994; Smith, Lindsay, & Pryke, 2001). However, when eyewitnesses are exposed to biased lineups, the value of postdictors such as confidence and decision time is undermined (Charman, Wells, & Joy, 2011; Key et al., 2017).

We tested whether general and eyewitness-specific self-ratings of memory ability can be used to discriminate identification performance, based on theoretical frameworks of metamemory. In particular, we aimed to investigate the efficacy of metamemory factors as postdictors of eyewitness identification for biased and unbiased lineups. Metamemory refers to the knowledge and awareness that an individual has about his or her own memory capabilities (Dunlosky & Bjork, 2008). This introspective knowledge is often used to monitor and control one's own memory performance. Research on metacognitive judgements has expanded rapidly, focusing on how well people think they have learned new information (i.e., judgements of learning; Double, Birney, & Walker, 2018) and how well people feel they recognize a particular information (i.e., feeling of knowing; Koriat, 2000). One predominant view is that metacognitive judgments are inferential in nature, involving a variety of heuristics and cues that have some degree of validity in predicting objective memory performance (Koriat, Ma'ayan, & Nussinson, 2006). Cues that can influence metamemory judgements can be divided into experience-based (the subjective learning experience) or

information-based (people's beliefs about their own memory capacities and limitations; Koriat, Nussinson, Bless, & Shaked, 2008). For example, metamemory judgements can be influenced by how quickly or easily an item is processed or accessed (Frank & Kuhlmann, 2017) and by preconceived notions about one's own competence in the domain tested (Dunning, Johnson, Ehrlinger, & Kruger, 2003). In forensic settings, eyewitnesses may produce confidence statements or identification decisions that are partially based on intrinsic cues of self-capacity (Leippe & Eisenstadt, 2014)

Brewer and Sampaio (2012) argue that confidence judgements result from the integration of two key components: information related to products and processes of the memory, and the individual's metamemory beliefs. In this prediction, confidence judgements are based partly on the learning experience, and partly on domain-specific beliefs (e.g., "My memory is not so good"; Hertzog & Dixon, 1994). Studies investigating the role of domain-specific beliefs in eyewitness confidence reports are sparse. Olsson and Juslin (1999), for example, found that individuals who considered themselves to be good face recognizers were more accurate and had a stronger confidence-accuracy relation in lineup identifications. Similarly, Perfect (2004) found that self-rated ability in the domain of eyewitness memory (i.e., face recognition and episodic details) was predictive of confidence judgements in a cued-recall task. These initial findings suggest that expressions of confidence in eyewitness settings may be influenced by witnesses' beliefs about their own memory ability.

Some longitudinal studies have shown a positive relation between memory self-efficacy and memory performance in different tasks (Seeman, McAvay, Merrill, Albert, & Rodin, 1996; Valentijn et al., 2006). Regarding face recognition ability, different tests of subjective and objective performance have been proposed as postdictors of identification accuracy and proclivity to choose (Bindemann, Brown, Koyas, & Russ, 2012; Russ et al., 2018). In the face matching literature, moderate to large correlations between self-reported

face perception ability and performance in objective face matching tests have been documented (Gray, Bird, & Cook, 2017; Shah, Sowden, Gaule, Catmur, & Bird, 2015; Ventura, Livingston, & Shah, 2018). However, studies focusing specifically on face recognition tasks have shown that individuals have limited insight into their ability to recognize unfamiliar faces (Bobak, Mileva, & Hancock, 2018). It has been argued that individuals tend to overgeneralize their ability to recognize familiar faces to situations in which unfamiliar faces need to be identified (Bindemann et al., 2014). Therefore, it might be expected that the association between self-ratings of memory capacity and objective memory functioning should be strongest when the self-rated ability is specific to the targeted memory task.

Contemporary memory models propose that memory consists of relatively independent systems (Baddeley, 2000; Tulving, 2007). Different memory systems can share some basic features (e.g., the means of acquiring new information), but they differ in some other features (e.g., functions, operating principles, and underlying neural mechanisms; Schacter, Wagner, & Buckner, 2000; Tulving, 2007). Thus, it can be expected that perceived lack of ability in one domain (e.g., semantic memory) may not be predictive of perceived failure in an eyewitness-relevant domain (e.g., memory of faces or episodic memory). Metamemory questionnaires often assess self-perceived performance and functioning of general memory – for example, dimensions of memory ability in the Multifactorial Metamemory Questionnaire (MMQ; Troyer & Rich, 2002). An instrument more relevant for the current research is the Eyewitness Metamemory Scale, which measures self-rated ability and endorsement of strategies for face and person recognition (Saraiva et al., 2019). In the current study, we tested both self-ratings of general memory ability and self-ratings of eyewitness memory ability as predictors of lineup identification performance in both biased and unbiased lineups.

A lineup can be considered biased when the suspect differs noticeably from other lineup members so that the suspect ‘stands out’ among the lineup options (Wells et al., 1998). In such instance the lineup fillers are implausible, and do not serve as functional alternatives to the suspect (Tredoux, 1999). It is typically observed that biased lineups produce more guilty and innocent suspect identifications, and fewer filler identifications than unbiased lineups (Fitzgerald, Price, Oriet, & Charman, 2013). Filler selection is thus a balancing act. On the one hand, increasing similarity between suspect and fillers increases the diagnosticity of suspect identifications. On the other hand, if fillers are too similar to the suspect the lineup may be too difficult for witnesses to make accurate identifications (Fitzgerald et al., 2013; Tredoux, 2002). In practice, some guidelines recommend that in order to protect the potentially innocent suspect from false identification, the suspect should not stand out in the lineup by being physically different from the fillers (Technical Working Group for Eyewitness Evidence, 2003). Given the difficulties in producing fillers that are appropriately similar to the suspect or to a description of the suspect, many lineups in the field are not fairly constructed (Memon et al., 2011).

Another issue with biased lineups is that they undermine the effectiveness of postdictors of accuracy such as identification confidence and decision time (Charman et al., 2011; Key et al., 2017). That is because subjective likelihood judgments are often based on comparisons between the chosen option and each of the individual alternatives. If implausible alternatives are present, there is increased perceived support for the chosen option, consequently inflating confidence judgements (Windschitl & Chambers, 2004). Charman et al. (2011) found that the presence of highly dissimilar fillers inflates witnesses’ confidence in mistaken identifications. Similarly, Key et al. (2017) suggested that when the suspect stands out, witnesses tend to be overconfident and faster (regardless of accuracy) compared to witnesses exposed to unbiased lineups. Taken together, these findings suggest that confidence

and decision time, normally effective postdictors of identification accuracy, have little diagnostic value if the identification decision was made from a biased lineup.

Our predictions about the relationship between lineup fairness and self-assessments of memory capacity draw from the literature on metamemory and task difficulty. In unbiased lineups, eyewitnesses need to rely more on their memory trace of the perpetrator to recognize one of the lineup members as a match of the remembered suspect's appearance. In contrast, biased lineups may be perceived as easier because fillers are less similar to the suspect and therefore are implausible options. This perceived lower difficulty in biased lineups creates a potentially misleading heuristic for metamemory judgements based on perceptual fluency. It has long been known that manipulations of perceptual fluency during retrieval can produce memory illusions (Jacoby & Whitehouse, 1989). One example is the belief that a more easily perceived test item is likely to be an old item. As such, memory misattribution may occur if perceptual ease is mistakenly assumed to indicate the stimulus's prior presentation (Higham & Vokey, 2000). In fact, under conditions of perceptual ease, metacognitive calibration tends to be weak, erring on the side of overconfidence (Chandler, 1994). Therefore, it might be expected that – if self-ratings of memory capacity are related to identification performance – this relation will be weaker for biased compared to unbiased lineups.

The current study

The purpose of this study was to test general and eyewitness-specific self-ratings of memory capacity as predictors of eyewitness identification performance, for both biased and unbiased lineups. We hypothesized self-ratings of memory capacity to be related to eyewitness identification accuracy (H1) and that this relationship would be stronger for self-capacity in eyewitness-specific memory domains compared to self-capacity in general memory domains (H2). Furthermore, we predicted individuals with higher self-ratings in the

metamemory factors to display have a stronger confidence-accuracy calibration than individuals with lower ratings (Olsson & Juslin, 1999; H3). Finally, we expected the relation between metamemory factors and eyewitness identification performance to be weaker for biased than unbiased lineups (H4).

Method

The present study was pre-registered and approved via the Open Science Framework (https://osf.io/ymkz9/?view_only=49c11c762050470fbe45880af51512ee).

Participants

A total of 1103 participants completed the study. We applied several exclusion criteria to ensure data quality: (a) 34 cases were removed for taking more than 90 minutes to complete the experiment, and (b) 97 cases for completing the experiment in under 15 minutes; (c) 95 cases were removed for not passing at least 4 out of 5 attention checks; and (d) 44 cases were removed due to suspicious bot activity (Prims & Motyl, 2018). The final sample ($N = 744$) had a mean age of $M = 29.98$, ranging from 18 to 72 years ($SD = 12.63$) and was comprised of 63% female participants. Most participants were workers from Amazon Mechanical Turk (54%), followed by university students (34%) and participants found through social media (12%). Participants recruited via Amazon Mechanical Turk received U\$1.00, students received course credits, and participants from social media were entered into a prize drawing for two £50 Amazon vouchers.

Our sample size was determined based on the confidence-accuracy calibration analysis, given that it is the most demanding analysis in our design. There are no clear guidelines on sample size requirements for calibration analysis, so we evaluated previous studies and reasoned that 400 choosers would provide stable estimates for calibration curves with five confidence levels (Sauer, Brewer, Zweck, & Weber, 2010; Sauerland & Sporer,

2009). Each participant completed two identifications so our total amount of observations was 1488, of which 815 were choosers.

Materials and Instruments

Eyewitness Metamemory Scale (EMS). The EMS contains 23 items divided into three factors: Contentment, Discontentment, and Strategies (Saraiva et al., 2019). All items are rated on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). The EMS-Contentment factor comprises 10 items (e.g., “My ability to remember faces is much better than other people’s ability to remember faces”; $\alpha = 0.85$) with higher scores indicating higher levels of memory contentment with respect to the ability to remember or recognise faces. The EMS-Discontentment factor has eight items (e.g., “Sometimes I have trouble recognizing a person that I know relatively well”; $\alpha = 0.89$) with higher scores indicating higher memory discontentment with ability to remember or recognise faces. The EMS-Strategies factor comprises five items (e.g., “I often create a visual image in my mind of a face that I want to remember”; $\alpha = 0.81$) with higher scores indicating higher endorsement of memory strategies to remember faces.

General metamemory instruments. In addition to the EMS, participants also completed the Multifactorial Memory Questionnaire (Troyer & Rich, 2002), and the Squire Subjective Memory Questionnaire (Bergen, Brands, & Jelicic, 2010; Squire, Wetzel, & Slater, 1979). The MMQ consists of three factors: Contentment ($\alpha = .92$), Ability ($\alpha = 0.92$), and Strategy ($\alpha = 0.88$). The contentment factor has 18 items (e.g., “I am generally pleased with my memory ability”) rated from 1 (*strongly agree*) to 5 (*strongly disagree*), with higher scores indicating higher memory contentment. The ability factor has 20 items related to experiences with common memory errors over the past two weeks (e.g., “How often do you forget an appointment?”) from 1 (*all the time*) to 5 (*never*), with higher scores indicating

better self-reported ability. The strategy factor has 19 items concerning the use of memory strategies during the past two weeks (e.g., “How often do you use a timer or alarm to remind you when to do something?”). The items are assessed on a scale ranging from 1 (*never*) to 5 (*all the time*), with higher scores indicating greater use of memory strategies. The SSMQ consists of 18 items related to the development of memory functioning (e.g., “My ability to recall things when I really try is”), rated on a 9-point scale ranging from -4 (*worse than ever*) to 4 (*better than ever before*).

Stimulus event. Participants viewed a 75s film depicting a thief stealing a phone from a victim (adapted from Sauerland, Wolfs, Crans, & Verschuere, 2017, Experiment 4). There were two versions of the video counterbalancing the role of two actresses (victim and perpetrator) to better generalize the results to different suspects. Thus, in one version actress A was the perpetrator, while in the other version actress B was the perpetrator.

Lineups. Every participant received two lineups, one for the perpetrator and one for the victim in the stimulus event. All lineups were presented in a simultaneous format and could be either target-present or target-absent. Target-present lineups consisted of five fillers and the target (i.e., victim or perpetrator) and target-absent lineups consisted of six fillers. Target-presence and the position of each member in the lineup were randomized for every lineup presentation. Pilot tests were conducted to construct fair and unfair lineups. In those tests, participants read a description of the target and were asked to select the person who best matched this description from a lineup of six members. Tredoux’s E was used as a measure of lineup fairness (Tredoux, 1998). Tredoux’s E takes a minimum value of 1 and a maximum value that equals the nominal lineup size (six in this case). If some lineup members are selected less often than expected by chance, E values decrease toward 1 depending on the number of lineup members falling below chance levels of choosing. Four pilot tests were conducted with a total of 123 participants, adapting the lineups to create sufficiently fair and

unfair lineups. The final four fair lineups (i.e., target present and target absent for each of the two targets) had Tredoux' E values ranging from 3.81 and 4.57, while the four unfair lineups had Tredoux' E values ranging from 1.54 to 2.56.

Procedure

Participants were recruited to an online experiment presented via Qualtrics. First, participants completed the EMS, followed by the MMQ and SSMQ. The EMS was always shown first, while the MMQ and SSMQ were presented in random order. Participants then watched the mock crime film, followed by a 5-minute filler task. Next, the first lineup was presented and participants were asked to identify the target or choose a 'not-present' option, while also providing a confidence judgement on a scale that ranged from 0% (*not confident at all*) to 100% (*totally confident*). After a 5-minute filler task, participants received the second lineup. The order of lineup presentation was randomized for every participant (i.e., either perpetrator first or victim first). Finally, some demographic information including gender, age and educational level was requested.

Results

In our analysis we tested choosers and nonchoosers separately for two reasons. First, it has been documented that postdictors of identification performance have different associations for choosers vs nonchoosers (e.g., Sauerland & Sporer, 2007, 2019). Second, triers of fact are more specifically concerned with eyewitnesses who choose someone from a lineup, rather than eyewitnesses who reject a lineup (Mickes, 2015).

Metamemory as predictors of eyewitness identification accuracy

First, we focused on the relation between metamemory and eyewitness identification accuracy by fitting regression models with metamemory factors as predictors of eyewitness

identification accuracy for choosers and nonchoosers. Each participant made two lineup decisions. Thus, the data were nested in two levels, with identification trials at Level 1 and participants at Level 2. Accordingly, we first tested for the necessity of using mixed effects models in order to account for the nested components of the data. The lme4 package available for R was used for all multilevel modelling (Bates, Mächler, Bolker, & Walker, 2014).

Global models were fitted including all metamemory factors as predictors of each outcome variable (Burnham & Anderson, 2002). All predictors were centred around their grand mean, subtracting the overall mean of that variable from each subject's score. Across the different models, we found that the intraclass correlation coefficients for participants ranged from .00 to .11 and ICC for lineups were all 0.00. We further conducted likelihood ratio tests comparing random-intercept models and fixed-intercept models for each outcome variable and found no evidence that random-intercept models fit the data significantly better than the fixed-intercept models for all outcomes (see Table S4.1 in Appendix 3). Taken together, these results do not support the use of random coefficient modelling (Burnham & Anderson, 2002). Therefore, we proceeded with estimating logistic regression models with no random effects. Table 4.1 provides descriptive statistics and correlation among all metamemory scales. Correlations ranged from $r = -.55$ to $r = .61$. Two out of six diagnostic tests pointed to the presence of multicollinearity in the model, but inspection of variance inflation factor, tolerance, Farrar-Glauber F-tests, and partial correlations revealed negligible multicollinearity, so we proceeded without adopting remedial measures.

Table 4.1

Means, Standard Deviations, and Correlations of the Metamemory Factors

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6
1. EMS-Contentment	4.20	1.02						
2. EMS-Discontentment	3.53	1.09	-.32**					
3. EMS-Strategy	4.50	1.09	.41**	.05				
4. MMQ-Contentment	3.62	0.70	.30**	-.55**	.07**			
5. MMQ-Ability	3.57	0.63	.31**	-.38**	.17**	.55**		
6. MMQ-Strategy	2.86	0.64	.08**	.24**	.19**	-.31**	-.44**	
7. SSMQ	5.89	1.21	.61**	-.22**	.35**	.47**	.41**	.05*

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. *M* = Mean. *SD* = Standard deviation. EMS = Eyewitness Metamemory Scale. MMQ = Multifactorial Metamemory Questionnaire. SSMQ = Squire Subjective Memory Questionnaire.

We hypothesized that self-ratings of memory capacity would be related to eyewitness identification accuracy (H1) and that this relationship would be stronger for self-capacity in eyewitness-specific memory domains compared to self-capacity in general memory domains (H2). Our first set of model testing focused on choosers, fitting one model for biased lineups and another model for unbiased lineups (see Table 4.2). Among choosers, higher scores in EMS-Discontentment (i.e., memory discontentment with ability to remember or recognise faces) were indicative of lower accuracy for both biased ($OR = 0.57, p < .001$) and unbiased lineups ($OR = 0.56, p < .001$; see Figure 4.1). For biased lineups, choosers were also more likely to be inaccurate if they claimed their memory improved over time (higher SSMQ; $OR = 0.65, p = .005$), but this effect was not observed for unbiased lineups ($p = .83$). None of the other metamemory factors were significant predictors of choosers accuracy for biased and unbiased lineups. We then repeated the same steps for the nonchoosers subset, fitting logistic regression models using metamemory factors as predictors of accuracy for biased and

unbiased lineups (see Table 4.3). The results showed that higher scores in general memory contentment (MMQ-Contentment) among nonchoosers were indicative of higher accuracy in biased lineups. However, this effect was very close to our stipulated alpha level of .05 and may be a spurious result observed by chance ($OR = 1.43$ $p = .03$). None of the other metamemory factors were significant predictors of nonchoosers identification accuracy for both biased and unbiased lineups.

Table 4.2

Logistic Regression Models of Metamemory Factors as Predictors of Identification Accuracy Among Choosers

Predictor	Biased Lineups			Unbiased Lineups		
	<i>B</i> (SE)	<i>p</i>	OR [95% CI]	<i>B</i> (SE)	<i>p</i>	OR [95% CI]
EMS-Contentment	0.05 (0.13)	.69	1.05 [0.80, 1.37]	-0.19 (0.15)	.22	0.83 [0.61, 1.11]
EMS-Discontentment	-0.55 (0.14)	<.001***	0.57 [0.43, 0.75]	-0.57 (0.14)	<.001***	0.56 [0.42, 0.74]
EMS-Strategies	0.08 (0.12)	.53	1.08 [0.84, 1.39]	-0.17 (0.12)	.18	0.84 [0.65, 1.08]
MMQ-Contentment	0.23 (0.16)	.14	1.26 [0.92, 1.72]	-0.16 (0.15)	.27	0.85 [0.63, 1.13]
MMQ-Ability	0.08 (0.15)	.56	1.09 [0.81, 1.45]	-0.03 (0.15)	.83	0.96 [0.72, 1.30]
MMQ-Strategy	0.21 (0.13)	.11	1.23 [0.95, 1.60]	0.19 (0.13)	.15	1.21 [0.93, 1.57]
SSMQ	-0.43 (0.15)	.005**	0.65 [0.47, 0.87]	0.03 (0.16)	.83	1.03 [0.75, 1.42]

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. OR = Odds ratio. EMS = Eyewitness Metamemory Scale. MMQ = Multifactorial Metamemory Questionnaire. SSMQ = Squire Subjective Memory Questionnaire.

Table 4.3

Logistic Regression Models of Metamemory Factors as Predictors of Identification Accuracy Among Nonchoosers

Predictor	Biased Lineups			Unbiased Lineups		
	<i>B</i> (SE)	<i>p</i>	OR [95% CI]	<i>B</i> (SE)	<i>p</i>	OR [95% CI]
EMS-Contentment	-0.23 (0.18)	.20	0.79 [0.55, 1.13]	-0.21 (0.17)	.24	0.81 [0.57, 1.14]
EMS-Discontentment	-0.26 (0.16)	.10	0.77 [0.56, 1.05]	-0.31 (0.16)	.06	0.73 [0.53, 1.01]
EMS-Strategies	0.06 (0.14)	.65	1.06 [0.80, 1.41]	0.09 (0.13)	.47	1.10 [0.84, 1.43]
MMQ-Contentment	0.36 (0.16)	.03*	1.43 [1.03, 1.99]	-0.20 (0.19)	.29	0.81 [0.55, 1.19]
MMQ-Ability	-0.12 (0.15)	.42	0.88 [0.65, 1.20]	0.09 (0.17)	.60	1.09 [0.77, 1.54]
MMQ-Strategy	0.05 (0.14)	.69	1.06 [0.79, 1.41]	0.13 (0.16)	.38	1.14 [0.84, 1.56]
SSMQ	-0.09 (0.17)	.60	0.91 [0.64, 1.28]	0.11 (0.18)	.55	1.11 [0.78, 1.59]

Note. * $p < .05$. OR = Odds ratio. EMS = Eyewitness Metamemory Scale. MMQ = Multifactorial Metamemory Questionnaire. SSMQ = Squire Subjective Memory Questionnaire.

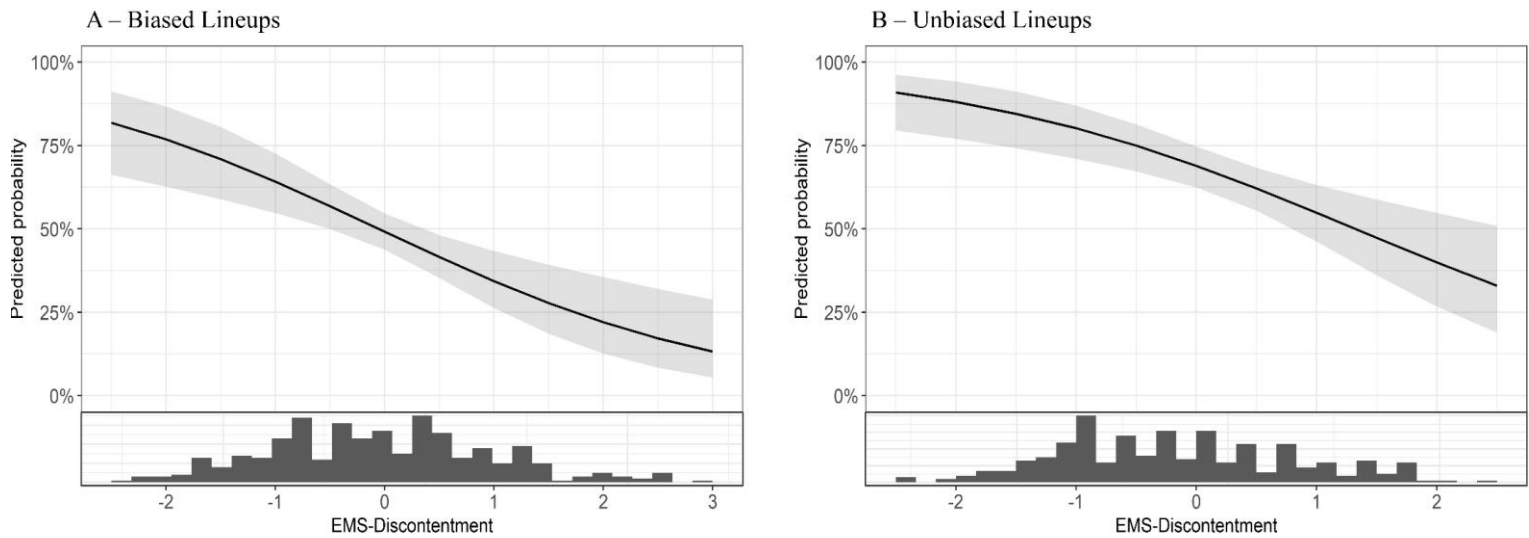


Figure 4.1. Predicted probabilities of identification accuracy among choosers as a function of EMS-Discontentment for biased and unbiased lineups. The shaded polygon represents 95% confidence intervals.

Metamemory and confidence-accuracy calibration

Calibration analyses were carried out to examine the relation between metamemory measures and the confidence-accuracy relationship in identification tasks. Following Brewer and Wells (2006), calibration curves were created by plotting the proportion of correct responses against five categories of confidence (0-20%, 30-40%, 50-60%, 70-80%, and 90-100%). We first produced calibration curves for choosers vs nonchoosers, and biased lineups vs unbiased lineups (see Figure 4.2). The diagonal line represents perfect calibration, such that each level of confidence is equivalent to the level of accuracy for decisions made with that level of confidence. Observations above this line indicate underconfidence, and observations below this line indicate overconfidence. We computed three calibration statistics: calibration index, over/underconfidence and resolution. Calibration (C) represents how far a given calibration curve is from a perfect calibration. It ranges from 0 (perfect calibration) to 1, and lower values represent better calibration. Over/underconfidence (O/U)

indicate if a curve strays more above or below the perfect calibration line, with values ranging from -1 (very underconfident) to 1 (very overconfident). The Normalized Resolution Index (*NRI*) represents how well confidence discriminates accurate from inaccurate identifications, with higher values indicating better discrimination (see Brewer & Wells, 2006). Following Palmer, Brewer, Weber, and Nagesh (2013) we used a jackknife procedure to compute standard errors for each calibration statistic, which were then converted to 95% inferential confidence intervals (Tryon & Lewis, 2008). If the confidence intervals do not overlap, that represents a significant difference (see Table 4.4). For choosers, the resolution statistic showed a high capability to discriminate between accurate and inaccurate identification decisions, for both biased (*NRI* = 0.12) and unbiased lineups (*NRI* = 0.11). However, choosers tended to be more overconfident in unbiased lineups (*O/U* = 0.14) compared to biased lineups (*O/U* = 0.03).

Table 4.4

Calibration Statistics for Choosers and Nonchoosers Across Biased and Unbiased Conditions

	C [95% CI]	O/U [95% CI]	NRI [95% CI]
Biased Lineups			
Choosers	0.01 [0.01, 0.02]	0.03 [-0.01, 0.07]	0.12 [0.06, 0.18]
Nonchoosers	0.03 [0.01, 0.04]	0.04 [-0.01, 0.09]	0.02 [-0.01, 0.04]
Unbiased Lineups			
Choosers	0.03 [0.01, 0.04]	0.14 [0.09, 0.18]	0.11 [0.05, 0.17]
Nonchoosers	0.03 [0.01, 0.04]	-0.03 [-0.08, 0.01]	0.04 [-0.01, 0.09]

Note. C = Calibration index. O/U = Over/underconfidence index. NRI = Normalized resolution index.

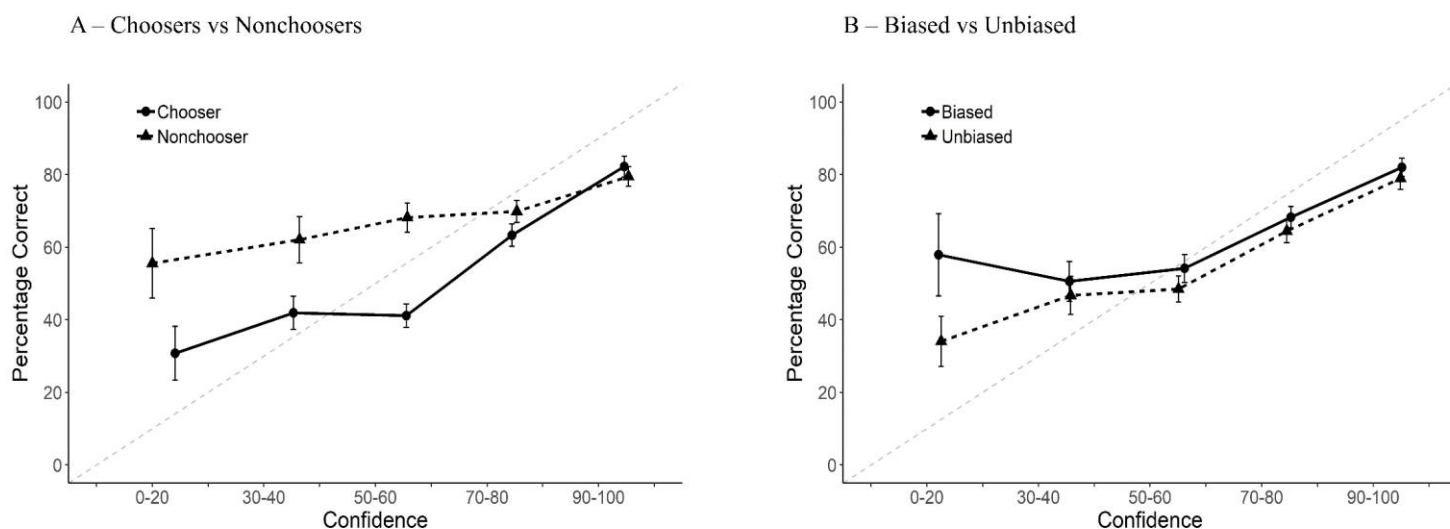


Figure 4.2. Confidence-accuracy calibration curves comparing choosers and nonchoosers (A) and biased vs unbiased lineups (B). The dotted diagonal grey line represents perfect calibration.

Next, we compared calibration statistics between high and low scorers on each of the metamemory measures. Following Olsson and Juslin (1999), individuals above the 66th percentile were selected as high scorers and individuals below the 33th percentile as low scorers. For this analysis, we focus on choosers, because triers of fact are more specifically concerned with eyewitnesses that choose someone from a lineup, rather than eyewitnesses that reject a lineup (Mickes, 2015; Wixted & Wells, 2017). Each metamemory group (low scorers and high scorers) had a mean sample size of $n = 271$. Inspection of the confidence intervals suggested that low scorers in the EMS-Contentment, EMS-Discontentment, EMS-Strategies, and SSMQ were significantly less overconfident than higher scorers in those components (see Figure 4.3). The calibration curves for those measures reveal that lower scorers were generally better calibrated than high scorers, especially for higher levels of confidence (see Figure 4.4). A similar pattern of results was observed for both biased and unbiased lineups (see Appendix 3).

Metamemory and Biased vs Unbiased lineups

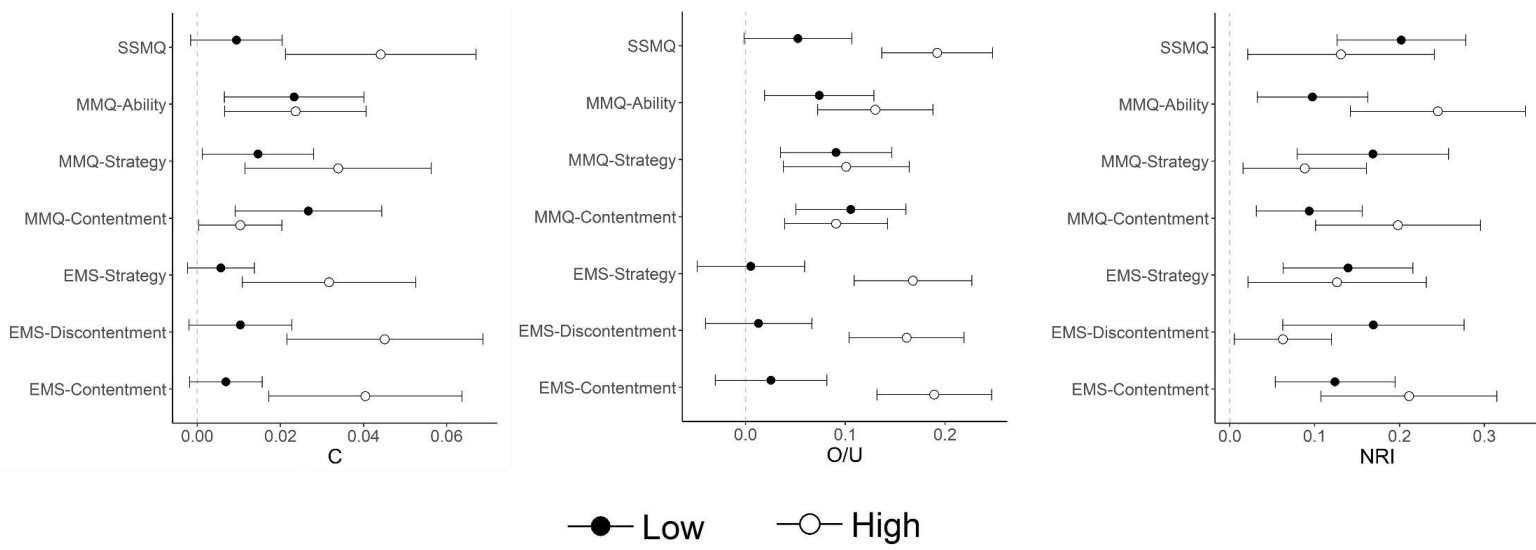


Figure 4.3. Inferential confidence intervals of calibration statistics for high and low scorers in each metamemory measure.

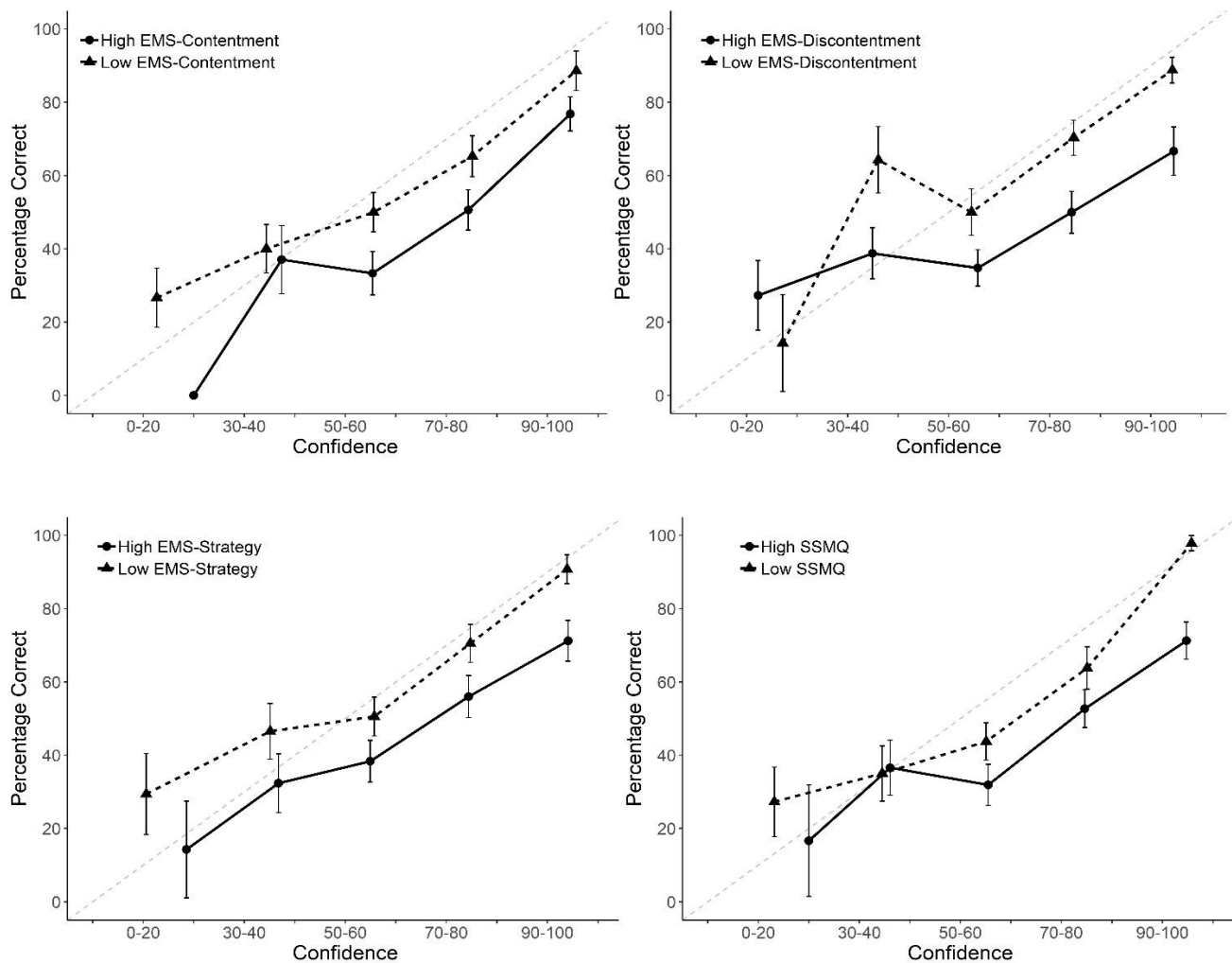


Figure 4.4. Calibration curves comparing low and high scorers in the EMS-Contentment, EMS-Discontentment, EMS-Strategy and SSMQ metacognitive factors.

Discussion

We investigated the diagnostic value of self-ratings of memory capacity on eyewitness identification accuracy and confidence, examining the relationship between memory self-capacity and identification performance for both biased and unbiased lineups. Our results revealed four key findings. First, higher discontentment with face recognition and person identification ability (EMS-Discontentment) was indicative of more inaccurate identifications for choosers in both biased and unbiased lineups. Second, in biased lineups, choosers were more likely to be inaccurate if they claimed their memory has improved over time (higher SSMQ). Third, among nonchoosers, higher scores in general memory

contentment were related to higher accuracy in biased lineups. Fourth, low scorers in EMS-Contentment, EMS-Discontentment, EMS-Strategies, and SSMQ were less overconfident and were generally better calibrated than high scorers, especially for higher levels of confidence. These findings contribute to an ongoing debate concerning the relationship between behavioural and self-reported face recognition ability. While some research suggests that individuals have only limited insight into their own face-recognition ability (Bindemann et al., 2014; Bobak et al., 2018), other studies report that self-ratings of face recognition ability are moderately to strongly related to objective performance (Livingston & Shah, 2018; Ventura et al., 2018).

Focusing specifically on eyewitness-identification paradigms, the current research provides initial evidence for a relation between self-reported memory discontentment and accuracy in lineup identification settings. Most notably, we expected the relation between self-ratings of memory ability and identification performance to be weaker in biased lineups compared to unbiased lineups, but this relation was similar for both conditions. This finding has important implications given that other postdictors of eyewitness identification performance are undermined in identifications made on biased lineups (Charman et al., 2011; Key et al., 2017). In other words, although confidence and decision time have reduced diagnostic value of accuracy in biased lineups, the same is not true for self-ratings of eyewitness memory ability. Charman et al. (2011) demonstrate that biased lineups reduce the diagnostic value of confidence because confidence is inflated when the lineup target is compared with implausible fillers. The authors also propose a scaling effect explanation for this finding, based on the fact that witnesses must generate anchor points when providing a similarity score between two faces on a subjective scale (such as a 1 to 7 scale). During an identification these anchor points may be affected by external factors, such as the dissimilarity between fillers and the target. In the case of self-ratings of memory ability, it is

less likely that broader ratings (i.e., “how good is your memory for faces?”) will be affected by situational factors such as filler dissimilarity. An individual that often distrusts his or her ability to recognize unfamiliar faces is unlikely to change this self-assessment when exposed to a biased or unbiased lineup. Therefore, specific self-ratings of eyewitness memory ability may be useful estimators of accuracy independently of lineup fairness. If replicated this finding may have important applied implications given the practical difficulties in producing unbiased lineups without computerized systems (Memon et al., 2011).

Another goal of the current study was to further investigate the relation between self-ratings of memory capacity and eyewitness confidence-accuracy relationship. Lower scores in all eyewitness metamemory factors (i.e., EMS-Contentment, EMS-Discontentment and EMS-Strategies) were indicative of a stronger confidence-accuracy relation among choosers, while higher scores in these factors were related to more overconfidence. This finding indicates that individuals who do not hold strong positive or negative opinions about their face recognition ability (low EMS-Contentment and low EMS-Discontentment) are more realistic when reporting their confidence, whilst individuals with a stronger opinion (i.e., either for low or high memory ability) tend to exaggerate their confidence assessments. Both individuals with low EMS-Contentment and low EMS Discontentment tended to be overconfident in their identifications. In contrast, Olsson and Juslin (1999) observed that individuals who rated themselves as good face recognizers had a more diagnostic confidence-accuracy relationship. However, in that study the authors acknowledge as a limitation having used single items of unknown validity and reliability, so inferences of memory self-efficacy from such a measure may be limited. The current data support the notion that individuals highly content with their own memories tend to exaggerate their confidence (Rickenbach, Agrigoroaei, & Lachman, 2015). The relation between higher discontentment and higher overconfidence seems less straightforward. One possible explanation for this result is that

choosers who are generally discontent with their own memories may overestimate their confidence precisely because they have selected someone from a lineup. In other words, if someone is discontent with their memory ability, but nevertheless select someone from a lineup, the selection may be followed by inflated confidence. Finally, individuals who claimed to endorse more memory strategies to encode faces were also more likely to be overconfident, possibly because those individuals feel that such strategies help them encode stronger memory traces (Chua, Hannula, & Ranganath, 2012). It is important to note, however, that score on the EMS-Strategies factor alone cannot inform whether participants used any strategies that they claimed to use.

Our prediction that eyewitness-specific metamemory factors would have a stronger relation to identification performance compared to general metamemory factors was only partially supported. We come to this conclusion because EMS-Discontentment was the strongest predictor in the models testing metamemory factors as predictors of identification accuracy among choosers. However, higher self-perceived general memory development (SSMQ) was also associated with more incorrect identifications for biased lineups. Additionally, contentment with general memory ability (MMQ-Contentment) was the only significant predictor of correct lineup rejections among nonchoosers, although this effect was not very pronounced. Taken together, this differential pattern of results support the notion of domain-specific memory self-efficacy, defined as an individual's appraisal of his or her usual ability in a given memory domain (Hertzog & Dixon, 1994). Our findings also suggest that assessments of self-capacity focused on eyewitness specific domains (e.g., face and person identification) are more valuable than assessments of *general* memory ability in distinguishing accurate from inaccurate identifications among choosers.

The current study has some limitations. First, although we tested two different targets in our eyewitness paradigm, we only used one mock-crime video. We reasoned that the

inclusion of multiple target events could generate noise and affect the power of our analyses, so it remains to be determined whether the current findings would replicate when assessing witness performance for different types of target events. Second, the metamemory assessment occurred prior to the lineup identification tasks. In planning our procedure, we reasoned that exposure to the identification tasks before the completion of the metamemory assessment would have affected self-ratings of memory ability to a greater extent than completing the assessments would affect eyewitness performance (Olsson & Juslin, 1999). This may have been appropriate for the aims in the current study, but future investigation should examine the robustness of self-rated memory ability as predictors of eyewitness performance when measures are obtained after the identification tasks.

Taken together, our findings contribute to the ongoing challenge of distinguishing accurate from inaccurate eyewitness identifications in the criminal justice system. We present initial evidence that choosers who report higher discontentment with their face and person identification ability are more likely to commit a false identification. Further work is necessary to determine the generalizability of these results to different target events and for metamemory assessments obtained after the identification tasks. Furthermore, metamemory assessments can increase the diagnostic value of confidence, given the observation that individuals with stronger opinions about their face recognition ability tend to be overconfident. This is of importance because confidence statements are often used to discriminate accurate from inaccurate witnesses, but little is known about whether confidence statements are affected by individual differences related to self-efficacy.

CHAPTER 5

Using general and eyewitness-specific metamemory assessments to estimate performance in multiple identifications

The research reported in this Chapter has been prepared for publication in: Saraiva, R. B., Van Boeijen, I. M., Hope, L., Horselenberg, R., Sauer, J., & Van Koppen, P (under review). Using general and eyewitness-specific metamemory assessments to estimate performance in multiple identifications.

Abstract

Identifications made by eyewitnesses are compelling evidence for prosecuting a suspect, but inaccurate identifications can have severe consequences including the conviction of innocent persons. Based on metacognitive frameworks, we aimed to distinguish accurate from inaccurate identification decisions by using metamemory assessments. Participants ($n = 203$) first completed an assessment of general and eyewitness-specific metamemory domains, followed by eight successive lineup identifications. We found that self-rated ability in the eyewitness memory domain was predictive of correct identifications, false identifications, confidence and confidence-accuracy calibration. Curiously, higher self-rated ability in the more general memory domain was predictive of fewer correct identifications. We discuss the potential applied value of metamemory assessments as predictors of correct and false lineup identifications, as well as theoretical contributions to underlying mechanisms of confidence.

Keywords: Eyewitness testimony; Memory; Metamemory; Face recognition; Identification; Lineup

Identifications are compelling evidence for prosecuting a suspect in criminal cases, but inaccurate identifications can have severe consequences including the conviction of innocent suspects (Wells & Olson, 2003). Distinguishing accurate from inaccurate eyewitnesses' identifications is thus an important but challenging issue in the criminal justice system. Much research has focused on factors that can postdict eyewitness identification accuracy, including mainly confidence statements (Brewer & Wells, 2006; Sporer, Penrod, Read, & Cutler, 1995), decision-time (Sauerland, Sagana, Sporer, & Wixted, 2018), and decision processes (Dunning & Stern, 1994; Sauerland & Sporer, 2007). More recently, tests of face recognition performance have also been proposed as informative estimators of identification accuracy and proclivity to choose (Baldassari, Kantner, & Lindsay, 2019; Russ, Sauerland, Lee, & Bindemann, 2018). However, very little work has focused on the relation between self-reported memory ability and eyewitness identification performance. In this study, we tested whether self-ratings of memory ability can be used to discriminate eyewitness's identification accuracy and confidence-accuracy realism, based on theoretical frameworks of metamemory.

Self-judgements about the accuracy of our memories are prevalent in many settings. A student taking a test with a guessing penalty must make that judgement, as must an interviewer relying on his or her notes, as must an eyewitness sworn to provide a truthful account. The study of *metamemory*, the knowledge about one's memory capabilities and about strategies that can aid memory, has expanded rapidly over the years (Dunlosky, Mueller, & Thiede, 2016; Koriat, Nussinson, Bless, & Shaked, 2008). Much of that work has focused on how beliefs about memory ability are formed and maintained (Hertzog & Dixon, 1994). Koriat et al. (2008) distinguish between two classes of cues that can influence metamemory judgments: experienced-based and information-based cues. Experience-based cues are intrinsic to the subjective learning experience, such as how quickly or easily an item

is processed or accessed (Frank & Kuhlmann, 2017). In contrast, information-based cues involve people's beliefs about their own memory capacities and limitations. When students are asked to judge their performance on a test, for example, their judgements may be based on preconceived notions about their competence in the domain tested (Dunning, Johnson, Ehrlinger, & Kruger, 2003). Likewise, eyewitnesses providing information about a crime (Perfect, 2004) or identifying suspects (Olsson & Juslin, 1999) produce confidence statements that are partially based on intrinsic cues of self-capacity (Leippe & Eisenstadt, 2014).

Considerable evidence supports the notion that eyewitness confidence has a positive, albeit not perfect, relationship with identification accuracy (Brewer & Wells, 2006; Wixted & Wells, 2017). That is because confidence judgements are malleable and may be contaminated by many different factors such as biased lineup instructions (Leippe, Eisenstadt, & Rauch, 2009), foil similarity (Charman, Wells, & Joy, 2011), and positive identification feedback (Wells & Bradfield, 1998). Intrinsic factors, such as memory self-efficacy, may also have an important role in eyewitness confidence statements (Leippe & Eisenstadt, 2014). Perfect (2004), for example, found that self-rated ability in eyewitness memory domain (i.e., face recognition and episodic details) was predictive of confidence judgements in a cued-recall task. Regarding identification tasks, Olsson and Juslin (1999), found that individuals who rated themselves as good face recognizers were more accurate and had a stronger confidence-accuracy relation in lineup identifications. This initial evidence suggests that self-ratings of memory ability may be useful to estimate confidence realism among eyewitnesses, however the literature on the subject is sparse.

Some evidence points to a positive relation between memory self-efficacy and memory performance in different tasks (Seeman, McAvay, Merrill, Albert, & Rodin, 1996; Valentijn et al., 2006). In the face matching literature, some studies show moderate to large

correlations between self-reported face perception ability and performance in face matching tests (Gray, Bird, & Cook, 2017; Ventura, Livingston, & Shah, 2018). However, studies focusing specifically on face recognition show that individuals have limited insight into their ability to recognize unfamiliar faces (Bindemann, Attard, & Johnston, 2014; Bobak, Mileva, & Hancock, 2018). Bindemann et al. (2014) argue that individuals tend to overgeneralize their ability to recognize familiar faces to situations in which unfamiliar faces need to be identified. It is arguable, then, that the association between memory self-efficacy and memory functioning is the strongest when the self-rated ability is specific to the targeted memory task.

Current theoretical perspectives posit that different memory systems share some basic features (e.g., the means of acquiring new information), and they differ in some other features (e.g., their functions and underlying neural mechanisms; Schacter, Wagner, & Buckner, 2000; Tulving, 2007). Metamemory questionnaires often assess self-perceived performance and functioning of general memory (Troyer & Rich, 2002). Other metamemory questionnaires focus on specific memory domains, such as the Self-evaluation of Memory Systems Questionnaire (Tonković & Vranić, 2011), which measures diverse systems such as episodic memory and semantic memory, and the Eyewitness Metamemory Scale, which measures self-rated ability and endorsement of strategies for face and person recognition (Saraiva et al., 2019). In the current study, we tested both self-ratings of general memory ability and self-ratings of eyewitness memory ability to estimate identification performance.

Some initial evidence has been found linking self-ratings of memory ability and eyewitness identification performance (Olsson & Juslin, 1999; Perfect, 2004). However, these studies measured memory self-capacity using items of untested reliability and validity. In the current study, we examined the relation between validated measures of metamemory and eyewitness identification performance. Additionally, in Olsson and Juslin (1999), the

confidence-accuracy relation could only be analysed at a group level (i.e., individuals with high performance versus individuals with lower performance), because confidence-accuracy relations could not be computed for each participant. In the current study, we investigated how overconfident or underconfident each participant was in a series of lineup identifications. In this approach, it is possible to compute confidence-accuracy relation scores that are specific to each participant, which may be tested as an outcome of metamemory measures. Multiple lineup identification designs have minimal effects on eyewitness accuracy, choosing and confidence (Mansour, Beaudry, & Lindsay, 2017). Hence, we adopted a multiple lineup identification design to allow for the calculation of calibration and over/underconfidence statistics for each participant instead of calculating only group coefficients. We expected that: (H1) metamemory measures would be predictive of identification performance; (H2) individuals with higher metamemory scores would be better calibrated in their confidence-accuracy relationship; (H3) the relationship between metamemory measures and eyewitness identification performance would be stronger for metamemory measures specifically related to eyewitness identification domains.

Method

Participants

To estimate the required sample size a power analysis was conducted on a pilot dataset of 56 participants. Pilot data were used instead of simulated data because it may offer more precise estimates of the variance and random coefficients present in nested data (Snijders, 2005). Simulations were conducted on multilevel models including metamemory factors as predictors of eyewitness performance scores (Green & MacLeod, 2016). The simulations showed that a sample of 200 participants would be required to detect an odds ratio of 1.30 with 80% power for each predictor. So, we set 200 as our target sample size. The

chosen effect size was not established based on previous studies due to lack of investigations on metamemory scores as predictors of eyewitness performance. Instead, it was chosen as a minimum effect size to be of practical relevance (Gelman & Carlin, 2014).

Participants were undergraduate students at universities in the United Kingdom and The Netherlands, who participated in exchange for course credit. A total of 221 participants completed the study, but two participants were excluded for not passing at least four out of the five attention checks, and sixteen other participants were excluded because technical issues were reported (e.g., films not loading properly). The final sample ($n = 203$) comprised of 81% female participants and had a mean age of $M = 20.22$, ranging from 18 to 58 years ($SD = 4.08$).

Materials and Instruments

Stimuli event. Participants watched eight mock crime films, developed in previous studies by Colloff and colleagues (Colloff, Wade, & Strange, 2016; Colloff, in preparation). Each film was 30 seconds long and depicted a simulated crime (e.g., burglary, car theft, laptop theft, drink spiking). All the films showed different individuals acting as perpetrators and victims. Every participant watched all eight films, presented in a random order.

Identification tasks. For each film, participants were instructed to try to identify the perpetrator in the mock-crime, in either a target-present or target-absent simultaneous lineup. In the target present lineups the target (perpetrator) was presented with five fillers, and in the target-absent lineups, six fillers were shown without the target. Target-presence was randomized for each film, but every participant saw an equal number of target-present and target-absent lineups (i.e. four of each presence type). None of the target or fillers had distinctive feature and all fillers were selected based on modal descriptions of the targets (see Colloff et al., 2016). Participants were instructed that the perpetrator may or may not be in the lineup, and they could evaluate the pictures for as long as needed to make a decision.

Metamemory assessment. Participants' self-ratings of memory ability were assessed using two different psychometric instruments: The Multifactorial Metamemory Questionnaire (MMQ, Troyer and Rich 2002) and the Eyewitness Metamemory Scale (EMS; Saraiva et al., 2019). The MMQ was developed to assess separate dimensions of subjective memory ratings, with 57 items divided into three sub-scales: Contentment, Ability, and Strategy. The contentment sub-scale comprises 18 items (e.g., "I am generally pleased with my memory ability"; $\alpha = 0.92$) rated from 1 (*strongly disagree*) to 5 (*strongly agree*), with higher scores indicating higher levels of memory contentment. The ability sub-scale comprises 20 items concerned with experiences with common memory errors over the past two weeks (e.g., "how often do you forget an appointment?"; $\alpha = 0.89$) from 1 (*all the time*) to 5 (*never*), with higher scores indicating higher levels of self-reported memory ability. The strategy sub-scale comprises 19 items focused on the use of memory strategies during the past two weeks (e.g., "how often do you use a timer or alarm to remind you when to do something?"; $\alpha = 0.82$). The items are assessed on a scale ranging from 1 (*never*) to 5 (*all the time*), with higher scores indicating more frequent use of memory strategies.

The EMS is an instrument comprising 23 items assessing three factors: Contentment, Discontentment and Strategies. All items are rated on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). The EMS-Contentment factor comprises 10 items (e.g., "My ability to remember faces is much better than other people's ability to remember faces"; $\alpha = 0.82$) with higher scores indicating higher levels memory contentment with respect to ability to remember or recognise faces. The EMS-Discontentment factor has eight items (e.g., "Sometimes I have trouble recognizing a person that I know relatively well"; $\alpha = 0.82$) with higher scores indicating higher memory discontentment with ability to remember or recognise faces. The EMS-Strategies factor comprises five items (e.g., "I often create a visual

image in my mind of a face that I want to remember.”; $\alpha = 0.75$) with higher scores indicating higher endorsement of memory strategies to remember faces.

Procedure. The study was conducted online using the Qualtrics platform. Participants first completed the EMS and MMQ instruments, and each instrument was presented in a randomised order. After completing the metamemory assessment, participants were instructed that they would watch a series of films and answer some questions about them. A filler task lasting one minute was presented after each mock crime film. Participants were then presented with either a target-present or target-absent lineup and were asked to try to identify the perpetrator in the film or choose a ‘not-present’ option. For each lineup, participants were asked to provide a confidence statement about their choice using a scale ranging from 0% - not at all certain to 100% - totally certain. This procedure was repeated for all eight mock-crimes, with the film appearing first, followed by a two-minute filler task and then the lineup identification task. The order of the films was randomized for each participant. Demographic information was requested at the end of the experiment.

Analysis approach. In our design, each participant made multiple lineup decisions. Thus, the data were nested in two levels, with identification trials at Level 1 and participants at Level 2. In a similar approach to that of Mansour et al. (2017), we fitted separate multilevel mixed-effects models to examine the metamemory measures as predictors of lineup decisions. The first model tested for correct identifications and included only the target-present lineups, comparing the number of correct identifications of the target (coded as 1) versus all other decisions (coded as 0). The second model included only the target-absent data, comparing the number of correct rejections (coded as 1) versus foil identifications (coded as 0). For the whole data, we compared the number of false identifications (coded as 1) versus all other decisions (coded as 0). The last multilevel model was fitted to predict confidence statements in the lineup decisions. All effects were estimated using maximum

likelihood, with logit link functions for models including categorical outcomes (e.g., accuracy) and link functions for models including continuous outcomes (i.e., confidence). The lme4 package available for R was used for all multilevel modelling (Bates, Mächler, Bolker, & Walker, 2014). Global models were fitted including all metamemory factors as predictors of each outcome variable (Burnham & Anderson, 2002). All predictors were centred around their grand mean, meaning that we subtracted the overall mean of that variable from each subject's score.

Calibration analysis was used to test our hypothesis that metamemory measures would be predictive of participants over/underconfidence. In the calibration approach, the aim is to verify if an eyewitness is more accurate about the information for which they show the greatest confidence. The confidence-accuracy relationship is computed across every level of confidence, typically in 10% increments (e.g., a scale that ranges from 0% - "not at all certain", to 100% - "totally certain"). A perfect calibration occurs when the groups confidence level equals the percentage of accurate answers for that group (e.g., witnesses who express 80% confidence are accurate in 80% of their identifications). Typically, a calibration index is computed, comprising the average squared discrepancy for each decision in a given confidence level and actual proportion of correct decisions in the same confidence group. A calibration index of 0 indicates perfect calibration, and, as such, values closer to 0 indicate a stronger confidence-accuracy relationship. Another related statistic is over/underconfidence (OU), which is computed by subtracting the mean accuracy from mean confidence for the entire witness sample, with scores that range from -1 (underconfidence) to +1 (overconfidence; see Brewer & Wells, 2006 for details on calculations of calibration scores). In this study, we computed calibration and OU scores for each participant using their multiple identifications and associated confidence ratings. This way, each participant had calibration scores that could be included as outcomes in regression models. Multilevel

models were not used for those analysis because the calibration scores are computed by aggregating values from all identification decisions made by the same participant. Thus, multiple regression models were used to analyse the metamemory measures as predictors of participants calibration and over/underconfidence. The Benjamini-Hochberg correction was applied to all p values from the regression models to account for multiple testing and false discovery rates (FDR; Benjamini & Hochberg, 1995). Emphasis is given to results with both raw p -values and FDR lower than .05, but results with significant raw p -values and FDR lower than .10 are also presented.

Results

Table 5.1 provides the means and standard deviations for all relevant lineup identification variables. Before testing our hypotheses, we calculated Intraclass Correlation Coefficients (ICC) to determine the extent to which our dependent variables varied among participants and lineup trials. Across the different models, we found that the ICC for participants ranged from .01 to .18 and ICC for lineups ranged from .06 to .12. We further conducted likelihood ratio tests comparing random-intercept models and fixed-intercept models for each outcome variable and found that random-intercept models fit the data significantly better than the equal-intercept models for all outcomes (see Table S5.1 in Appendix 4). Taken as a whole, these results support our use of random coefficient modelling (Burnham & Anderson, 2002). Table 5.2 provides descriptive statistics and correlation among all metamemory scales. Correlations ranged from $r = -.41$ to $r = .60$, and no indication of multicollinearity was observed (all VIF < 1.83).

Table 5.1

Main descriptive statistics for identification performance scores

	All Lineups		Lineup 1		Lineup 2		Lineup 3		Lineup 4		Lineup 5		Lineup 6		Lineup 7		Lineup 8	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Correct identifications	581	36	72	35	66	32	57	28	92	45	73	36	74	36	59	29	88	43
Correct rejections	521	32	49	24	70	34	69	34	43	21	46	23	77	38	74	36	93	46
False Identifications	386	24	74	36	43	21	37	18	57	28	81	40	31	15	47	23	16	08

Table 5.2

Means, standard deviations, and correlations of the metamemory measures

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5
1. EMS-Contentment	3.96	0.86					
2. EMS-Discontentment	2.96	1.05	-.27** [-.39, -.14]				
3. EMS-Strategy	4.40	1.10	.23** [.10, .36]	.04 [-.10, .18]			
4. MMQ-Contentment	3.69	0.69	.30** [.17, .42]	-.22** [-.35, -.08]	.01 [-.13, .15]		
5. MMQ-Ability	3.50	0.56	.30** [.17, .42]	-.20** [-.33, -.07]	.10 [-.04, .24]	.60** [.50, .68]	
6. MMQ-Strategy	2.91	0.54	.01 [-.13, .15]	.15* [.01, .28]	.26** [.13, .39]	-.29** [-.41, -.16]	-.41** [-.52, -.29]

Note. * indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$. Values in brackets are 95% confidence intervals

The raw p -values and adjusted p -values using Benjamini-Hochberg correction for all regression models are presented in Table 5.3. The odds ratio, 95% confidence intervals, and random parameter estimates for the models including metamemory scales as predictors of eyewitness categorical outcomes are presented in Table 5.4. As shown in Table 5.4 EMS-Contentment was the main predictor of eyewitness identification outcomes. For each unit increase in EMS-Contentment score, the odds of making a correct identification increased by a factor of 1.41, and the odds of making a false identification decreased by a factor of 0.79. Considering raw p -values lower than .05, EMS-Contentment was also an indicator of correct rejections (OR = 1.21, $p = .02$), but this effect was not significant after applying a correction for false detection rates (adjusted $p = .08$). Figure 5.1 presents the predictive probabilities for

correct identifications, correct rejections, and false identifications as a function of EMS-

Contentment, holding all other variables in the model at their mean value. MMQ-

Contentment was also a significant predictor of correct identifications, but each unit increase

Table 5.3

Raw p-values and adjusted p-values using Benjamini-Hochberg correction for all regression models

	Correct identifications		Correct rejections		False Identifications		Confidence		Calibration		Over/underconfidence	
	Raw <i>p</i>	BH correction	Raw <i>p</i>	BH correction	Raw <i>p</i>	BH correction	Raw <i>p</i>	BH correction	Raw <i>p</i>	BH correction	Raw <i>p</i>	BH correction
EMS-Contentment	<.001	.005	.02	.08	.002	.02	<.001	.005	.003	.03	.25	.50
EMS-Discontentment	.33	.57	.87	.94	.52	.72	.90	.94	.55	.73	.68	.86
EMS-Strategies	.42	.61	.03	.10	.02	.09	.02	.08	.76	.91	.03	.11
MMQ-Contentment	.005	.03	.38	.57	.11	.24	.004	.03	.08	.21	.69	.86
MMQ-Ability	.05	.16	.37	.57	.85	.94	.33	.57	.96	.96	.15	.32
MMQ-Strategy	.10	.25	.35	.57	.87	.94	.29	.56	.04	.11	.91	.94

in its score decreased the odds of correct identifications (OR = 0.74).

Table 5.4

Logistic multilevel models testing metamemory scores as predictors of eyewitness identification performance

Fixed coefficients	Correct identifications		Correct rejections		False Identifications		Confidence	
	OR	95% CI	OR	95% CI	OR	95% CI	β	95% CI
EMS-Contentment	1.41	1.17,	1.21	1.02,	0.79	0.68,	0.34	0.15,

		1.70**		1.44		0.91*		0.51**
EMS-Discontentment	0.91	0.77, 1.09	0.99	0.84, 1.16	1.05	0.88, 1.22	-0.01	-0.18, 0.16
EMS-Strategies	1.04	0.87, 1.25	0.82	0.69, 0.97	1.19	1.03, 1.38	0.20	0.03, 0.38
MMQ-Contentment	0.74	0.59, 0.91*	0.91	0.75, 1.10	1.15	0.97, 1.36	-0.30	-0.50, - 0.09*
MMQ-Ability	0.81	0.64, 1.01	1.12	0.91, 1.38	0.97	0.81, 1.16	0.13	-0.10 0.32
MMQ-Strategy	0.87	0.72, 1.06	1.10	0.92, 1.31	0.97	0.83, 1.14	-0.10	-0.29, 0.09
Random parameters	Variance	SD	Variance	SD	Variance	SD	Variance	SD
Level 2: Participants	0.05	0.23	0.02	0.13	0.19	0.44	0.93	0.96
Level 1: Lineup	0.29	0.54	0.45	0.67	0.37	0.61	0.30	0.55

Note. * indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$. All p -values in the regression models were adjusted using Benjamini-Hochberg false discovery rate

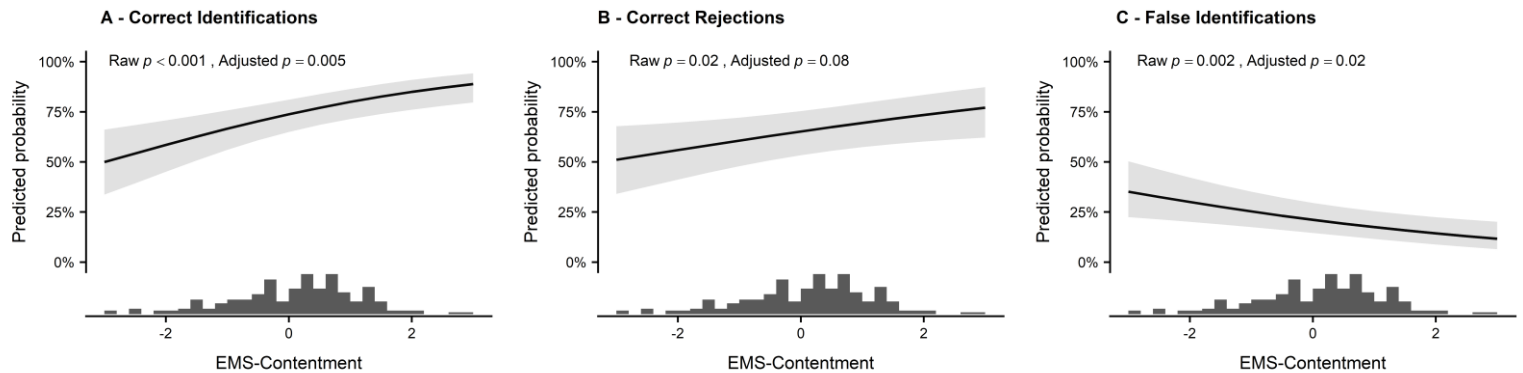


Figure 5.1. Multilevel regression plots of predicted probabilities for model outcomes as a function of EMS-Contentment. Regression lines are predicted relationships and shaded polygons represent 95% confidence intervals.

Considering non-adjusted p -values $< .05$, we found that EMS-Strategies was predictive of correct rejections and false identifications, with each unit increase in its score decreasing the odds of correct rejections by a factor of 0.82, and increasing the odds of false identifications by a factor of 1.19. To better examine this result, we explored a multilevel model including the metamemory scales as predictors of proclivity to choose in the lineup tasks. For this outcome, every target or filler identification was coded as ‘chooser’ and every lineup rejection was coded as ‘non-chooser’. The coefficients in the model revealed that

Table 5.5

Multiple regressions testing metamemory scores as predictors of calibration and over/underconfidence

	Calibration		Over/underconfidence	
Fixed coefficients	β	95% CI	β	95% CI
EMS-Contentment	-0.022	-0.038, -0.007*	-0.020	-0.055, 0.014
EMS-Discontentment	0.003	-0.008, 0.015	0.005	-0.021, 0.032
EMS-Strategies	-0.001	-0.013, 0.009	0.031	0.004, 0.057
MMQ-Contentment	0.019	-0.002, 0.040	0.009	-0.039, 0.059
MMQ-Ability	0.001	-0.027, 0.028	0.047	-0.018, 0.112
MMQ-Strategy	0.025	0.001, 0.052	0.003	-0.055, 0.061

Note. * indicates $p < .05$. All p-values in the regression models were adjusted using Benjamini-Hochberg false discovery rate

EMS-Strategies was the only significant predictor of choosing (see Table S5.2 in Appendix

4). With each unit increase in EMS-Strategies, the odds of selecting someone from a lineup (i.e., target or filler) increased by a factor of 1.16 (95% CI = 1.04, 1.30, $p = .005$).

The multilevel model including metamemory scores as predictors of confidence revealed that EMS-Contentment, EMS-Strategy, and MMQ-Contentment were predictive of confidence statements (see Table 5.3). Higher scores in EMS-Contentment ($\beta = 0.34$, adjusted $p = .005$) were indicative of higher confidence, and higher scores in MMQ-Contentment were indicative of lower confidence ($\beta = -0.30$, adjusted $p = .03$). Results from the multiple regression models including metamemory scores as predictors of participants confidence-accuracy calibration and overconfidence are presented in Table 5.5. Higher scores in EMS-Contentment were indicative of better calibration ($\beta = -0.02$).

Discussion

We tested the effectiveness of general and domain-specific metamemory as predictors of eyewitness identification performance. The findings indicated that self-rated memory ability for face and person identification, as measured by the EMS, was predictive of eyewitness identification performance. Specifically, higher EMS-Contentment scores were associated with more correct identifications and fewer false identifications. These findings are consistent with previous studies showing a relation between self-perceived memory

efficacy and memory performance (Kliegel & Jager, 2006; Valentijn et al., 2006). In contrast, Perfect (2004) observed that self-rated ability in the domains of eyewitness memory was not a useful predictor of eyewitness performance in identification and cued-recall tests. However, the self-efficacy measure used in that study was one single-item, so inferences of memory self-efficacy from such a measure may be limited. The current data support the notion that self-perceived memory ability for face and person identification contribute to the estimation of objective identification performance. This initial finding signals that assessments of self-memory ability may allow for more accurate predictions of correct and false lineup identifications, although replication and further research is necessary before further application is considered.

Surprisingly, we observed that higher scores on self-rated general memory capacity, as measured by the MMQ, was indicative of fewer correct identifications in target-present lineups. That seems to be the reverse of what was observed with the EMS-Contentment factor, which showed a positive relationship with correct identifications. It may be the case that individuals with higher scores on MMQ-Contentment are overestimating their general memory ability, which may lead to more false identifications and fewer correct identifications. This finding is in line with Rickenbach, Agrigoroaei, and Lachman (2015), who observed that self-assessments of episodic memory ability are often inaccurate relative to actual performance. These results suggest that while higher self-ratings of face memory capacity are indicative of more correct identifications, higher self-ratings of general memory ability are indicative of fewer correct identifications. In other words, individuals who claim to have a strong general memory ability, but weak face memory ability, may be at a higher risk of committing false identifications.

Self-rated ability for face memory and person identification was also related to confidence statements in identification tasks, with higher scores in EMS-Contentment being

associated to higher confidence. Koriat et al. (2008) argue that confidence judgements may be based partly on the learning experience, and partly on domain-specific beliefs (e.g., “My memory is not so good”). In our data, confidence judgements were predicted by self-rated ability for face memory and person identification tasks, lending support to the notion that confidence statements stem in part from self-rated ability in a domain of knowledge (Brewer & Sampaio, 2012). Furthermore, we found that individuals with higher scores in EMS-Contentment may be generally more effective at regulating their confidence judgments to reflect the likely accuracy of their identification decisions. This finding is consistent with those of Olsson and Juslin (1999), who initially found that individuals reporting higher self-rated recognition skill show a more diagnostic confidence-accuracy relation. Interestingly, confidence in lineup identifications was predicted only by self-rated ability for face and person identification, but not by factors on general memory ability. This finding supports the notion of domain-specific memory self-efficacy, defined as an individual’s appraisal of his or her usual ability in a given memory domain (Hertzog & Dixon, 1994).

Endorsement of memory strategies for face and person identification, as measured by the EMS-Strategies factor, was related to more false identifications and fewer correct rejections. These specific findings had a higher risk of false discovery after correcting for multiple testing, which means they may represent spurious effects found by chance. Nevertheless, we decided to discuss these results given the existence of some theoretical support for the observed pattern of results. Exploring this result, we found that EMS-Strategies was the only significant predictor of eyewitness choice, with higher scores in EMS-Strategies indicating a higher probability of choosing someone from a lineup. This finding contributes to the literature on individual differences and eyewitness’s proclivity to choose. Baldassari and colleagues (2019), for example, found that individual differences in proclivity to choose in a face recognition task was predictive of false identifications in

culprit-absent lineups. In fact, a growing literature suggests that individual differences related to performance on standardized tests may offer reliable indicators of eyewitness identification performance (Bindemann, Avetisyan, & Rakow, 2012; Russ et al., 2018). We expand on those findings by suggesting that individuals who endorse more strategies to better remember faces may have a higher proclivity to choose in lineup identifications, contributing to more false identifications and fewer correct rejections. It may be the case that individuals who claim to endorse more strategies may choose someone from a lineup more frequently because they feel that such strategies help them encode stronger memory traces (Chua, Hannula, & Ranganath, 2012). It is important to note, however, that score on the EMS-Strategies factor alone cannot inform whether participants used any strategies that they claimed to use.

There are some limitations associated with current study. First, the metamemory assessment occurred prior to the lineup identification tasks. In determining our procedure, we reasoned that exposure to the identification tasks before the completion of the metamemory assessment would have affected witnesses' self-ratings of their memory ability to a greater degree than completing the measures would affect eyewitness performance (Olsson & Juslin, 1999). This may have been appropriate for the aims in the current study, but future investigation should examine the robustness of self-rated memory ability as predictors of eyewitness performance when measures are obtained after the identification tasks. We also implemented an eyewitness paradigm with a short delay between exposure to the crime and the identification tasks. In naturalistic contexts, the delay between witnessing a crime and completing an identification task can take much longer (Valentine, Pickering, & Darling, 2003). It may be the case that self-ratings of memory ability are stable (Lane & Zelinski, 2003) and could still be predictive of identification accuracy in situations involving longer delays between encoding and recognition. However, it remains to be determined whether the

current findings would replicate when assessing witness performance for identifications made after longer delays.

Identifying factors that can distinguish accurate from inaccurate witnesses is an important challenge for the criminal justice system. In this paper we present initial evidence that self-rated memory ability for face and person identification may be a useful predictor of identification performance. Participants who reported higher contentment with their eyewitness memory ability had more correct identification, fewer false identifications, and were also better calibrated in their confidence-accuracy relation. If corroborated by further evidence, this finding could offer the basis for an assessment tool specifically tailored for eyewitness identification decisions. Such an assessment could be used in addition to other known predictors of identification accuracy, such as confidence statements, decision time, and decision processes, in order to better inform decisions regarding eyewitness identification evidence.

Data Availability Statement

All data and R data analysis script files can be found at:

https://osf.io/vu3ab/?view_only=26d1108258c8413e995320c7e0b0b084

CHAPTER 6

Using metamemory measures and memory tests to estimate eyewitness free recall performance

The research reported in this Chapter has been prepared for publication in:
Saraiva, R. B., Hope, L., Horselenberg, R., Ost, J., Sauer, J., & Van Koppen, P (under review). Using metamemory measures and memory tests to estimate eyewitness free recall performance

Abstract

Using a mock witness methodology, we investigated the predictive value of metamemory measures and objective memory tests as indicators of eyewitness free recall performance. Participants ($n = 208$) first completed a metamemory assessment that included assessments of self-rated memory capacity, memory development and use of strategies. In a separate session, participants watched a mock crime video and provided a free recall account, followed by one out of four independent memory tests (i.e., free recall, cued recall, general knowledge or face recognition). Accuracy, completeness, confidence and over/underconfidence in the eyewitness free recall were the main dependent variables. Results indicated three main findings: (1) subjective assessments of memory capacity were not related to eyewitness free recall performance; (2) individuals with higher self-rated memory capacity had a slightly stronger confidence-accuracy relation in free recall; and, (3) although individual confidence and over/underconfidence was somewhat stable across different memory tests, accuracy was less stable. These results are discussed with respect to metamemory assessments and performance stability across memory tests of different domains.

Keywords: Eyewitness testimony; metamemory; free recall; confidence; memory

Eyewitness statements are often critical in criminal investigations and may be the only available source of information about a crime when physical evidence is absent. Correct information provided by eyewitnesses can help investigators identify and trace suspects, but incorrect information can impede the investigative process, wasting valuable time and resources. Therefore, the ability to distinguish between accurate and inaccurate information is critical when evaluating witness's statements. Psychological research has identified some factors that can help discriminate eyewitness identification accuracy, such as confidence (Sauer, Brewer, Zuck, & Weber, 2010; Sporer, Penrod, Read, & Cutler, 1995; Wixted & Wells, 2017), decision time (Weber, Brewer, Wells, Semmler, & Keast, 2004), and decision processes (Sauerland & Sporer, 2007; Weber et al., 2004). However, only a small number of studies have investigated factors that may assist with the estimation of accuracy in free reports (Dahl, Allwood, Scimone, & Rennemark, 2015; Odnot, Wolters, & van Giezen, 2013; Sauer & Hope, 2016; Weber & Brewer, 2008). In the current research, we aimed to further investigate the relationship between accuracy and confidence in eyewitness free recall. We tested whether metamemory instruments and distinct memory tests can be used to estimate eyewitness accuracy, completeness, confidence and over/underconfidence in a free recall task pertaining to the witnessed event.

The Eyewitness Confidence-Accuracy Relationship

Confidence assessments may be present in many instances of criminal investigations, occurring whenever a police officer, lawyer or another law practitioner asks a witness if they are sure about a given account or identification. Lay people and practitioners in the criminal justice system often regard eyewitness confidence as a strong indicator of eyewitness credibility (Brewer & Burke, 2002; Penrod & Cutler, 1995). Many studies have shown that confidence can be a valid indicator of eyewitness identification performance, especially when

confidence statements are collected soon after identification decisions (Brewer, Keast, & Rishworth, 2002; Brewer & Wells, 2006; Sauer et al., 2010; Wixted, Mickes, Clark, Gronlund, & Roediger, 2015; Wixted & Wells, 2017). However, eyewitnesses may exaggerate or inflate their confidence in an identification, particularly if encoding conditions are bad or if biased lineup procedures are used, making confidence statements in such circumstances less reliable estimates of accuracy (Douglass & Jones, 2013; Leippe, Eisenstadt, & Rauch, 2009; Lindsay, Read, & Sharma, 1998; Penrod & Cutler, 1995)

Most research on eyewitness confidence-accuracy (CA) relationship has focused on suspect identifications, but witnesses' freely recalled memories are also highly relevant in forensic contexts, since most of the information provided by eyewitnesses comprise descriptions of the perpetrator and the event (Van Koppen & Lochun, 1997). Free recall paradigms usually generate a stronger CA relation when compared to forced-response or recognition memory paradigms (Robinson, Johnson, & Herndon, 1997; Robinson, Johnson, & Robertson, 2000; Robinson & Johnson, 1996). Studies focusing on free recall of staged crimes, for example, commonly find CA correlations of around .60 (Odinot & Wolters, 2006; Odinot et al., 2013; Robinson & Johnson, 1996). This stronger CA relation may be accounted for by the "free" component of recall tasks, because in such procedures witnesses tend to choose which information to report based on their confidence, consequently increasing metamemory realism (Allwood & Jonsson, 2005). In fact, theoretical frameworks of memory reporting suggest that people balance the demands for informativeness and accuracy during cued-recall, withholding details that fall below a pre-set criterion of probable accuracy (Koriat & Goldsmith, 1996; Robinson et al., 2000; Robinson & Johnson, 1996). However, important aspects of the CA relation in eyewitness free recall remain to be examined, particularly those related to individual differences in self-perceived memory capacity and memory functioning.

Metamemory And The CA Relationship

Metamemory refers to an individual's knowledge and awareness of his or her own memory capabilities, based on previous experiences and beliefs (Dunlosky & Bjork, 2008). This introspective knowledge is used to monitor and control encoding, retrieval, reporting processes, and to provide information about memory confidence (Koriat & Goldsmith, 1996; Nelson & Narens, 1994). During confidence assessments individuals may rely mostly on memory trace strength, typically providing higher confidence ratings to stronger memory traces (Koriat, Goldsmith, & Pansky, 2000). Notwithstanding, some authors argue that memory confidence may not only be related to memory strength, but also to intrinsic, heuristic and self-credibility cues (Jonsson & Allwood, 2003; Leippe & Eisenstadt, 2014). Intrinsic cues have been found to influence other types of metamemory judgements such as judgements of learning (Koriat, Bjork, Sheffer, & Bar, 2004) and feeling of knowing (Koriat, 2000). In legal settings, eyewitnesses might use the difficulty of the task as a cue to report confidence, showing less confidence for correct answers to hard questions when compared to correct answers to easy questions (Howie & Roebers, 2007; Stankov, 2000). Heuristic cues comprise one's beliefs about external factors that can help or impair memory encoding and retrieval, for example when eyewitnesses put more effort on recall tasks, even if it does not lead to changes in accuracy (Shaw & Zerr, 2003). These self-credibility cues can also be derived from people's beliefs about their overall memory performance. Some could overestimate their ability to recall events and show overconfidence, others may underestimate their memory ability and show underconfidence (Leippe, Eisenstadt, Rauch, & Stambush, 2006; Olsson & Juslin, 1999).

Memory self-credibility cues may have important implications not only on how individuals report confidence during free recall accounts, but also on the quantity and quality

of information disclosed. Evans and Fisher (2011) argue that eyewitness free recall reporting is, in its essence, a metacognitive control process in which individuals strategically withhold uncertain responses, or provide imprecise, but likely accurate responses (Koriat & Goldsmith, 1996). The trade-off between accuracy, quantity and precision of information poses the question of whether self-assessments of memory capacity relate to the completeness or accuracy of free recall reports.

Eyewitness Recall And Performance On Objective Memory Tests

Subjective self-ratings of memory capacity may be useful predictors of memory performance, but performance in objective memory tests might be more informative with respect to eyewitness accuracy, confidence and over/underconfidence. Some studies, for example, show that accuracy in different face memory tests are predictive of eyewitness identification accuracy (Bindemann, Brown, Koyas, & Russ, 2012; Morgan et al., 2007). However, to our knowledge, no research has examined the relation between eyewitness free recall performance and performance in unrelated memory tests, such as free-recall and cued-recall for a non-criminal event, or face recognition and general knowledge tests. On one hand, some evidence shows that different memory systems are rather independent (Melby-Lervåg & Hulme, 2013; Simons et al., 2016), suggesting that performance in one memory test may not be predictive of performance in other tests. On the other hand, a few studies indicate some stability in performance for different memory tests due to individual differences (Jaschinski & Wentura, 2002; Zhu et al., 2010). It may be expected that tasks more closely related to an eyewitness free recall paradigm (e.g., unrelated free recall or cued recall tests) are more predictive of eyewitness free recall performance than more distantly related tasks (e.g., face recognition and general knowledge). In the present experiment, we investigated the predictive utility of different objective memory tests (i.e., free recall, cued

recall, general knowledge and face recognition) for eyewitness free recall performance, in order to examine stability in accuracy, confidence, and over/underconfidence across different memory domains. A recognition test for unrelated faces was included as a test more distantly related to the eyewitness free recall, in order to better examine the extent to which memory performance stability may change depending on similarities between memory domains.

Current Research

The idea that confidence measures are affected not only by the availability of memory traces, but also by different intrinsic, heuristic and self-credibility cues is critical to the understanding and practical utility of confidence in forensic contexts. Previous studies have shown that self-assessments of memory capacity improve the diagnostic value of confidence in eyewitness identification tasks (Olsson & Juslin, 1999; Searcy, Bartlett, & Memon, 2000). In this study, we tested the use of metamemory self-assessment instruments, and objective memory tests as predictors of eyewitness accuracy, completeness, confidence and over/underconfidence in a free recall task. We hypothesized that self-reported metamemory measures would be predictive of eyewitness accuracy (H1), completeness (H2), confidence (H3); and over/underconfidence (H4). These predictions are based on some initial evidence that metamemory self-assessments are related to identification confidence, although mixed results can be found regarding the relation between metamemory and identification accuracy (Olsson & Juslin, 1999; Searcy et al., 2000).

Estimating eyewitness confidence using metamemory components relates to an important theoretical question, concerning the role of intrinsic cues on reports of confidence during free recall tasks. It was expected that individuals with higher metamemory scores would have a stronger confidence-accuracy relation than individuals with lower metamemory scores (H5). Finally, it was hypothesized that performance in different objective memory

tests would be predictive of eyewitness free recall performance (H6), and tests more closely related to the free recall test would be more predictive than tasks distantly related to this paradigm (H7).

Method

Participants And Design

A sample of 208 participants was recruited from local and student community (81% female; age between 18 and 70 years old, M age = 23.25, SD = 9.33). The required sample size was estimated using power analysis conducted for a Multivariate Regression Analysis with $f = 0.05$, $\alpha = .05$ and power = 0.95. The projected sample size needed for this effect size was approximately $N = 205$. Participants either received course credits or a £5 compensation for their time.

All participants completed a metamemory assessment and a free recall test of a mock-crime video. In a between-subjects design, participants were assigned to one of four possible objective memory test conditions: free recall, cued recall, face recognition, and general knowledge. The dependent variables were accuracy, completeness, confidence and over/underconfidence in the eyewitness free recall task.

Materials

Multifactorial Metamemory Questionnaire (MMQ; Troyer & Rich, 2002). The MMQ is an instrument with 57 items comprising three distinct factors: Contentment (i.e., affect related to memory abilities, $\alpha = 0.91$), Ability (i.e., frequency of memory problems in different situations, $\alpha = 0.89$), and Strategy (i.e., use of memory strategies in everyday life, $\alpha = 0.84$). The factor Contentment consists of 18 items (e.g., 'my memory is worse than most other people my age') rated on a scale from 1 (*strongly agree*) to 5 (*strongly disagree*), with

higher scores indicating higher memory contentment. The factor Ability has 20 items which requires respondents to indicate how often they experienced memory mistakes over the last two weeks (e.g., ‘how often do you forget an appointment?’), on a scale ranging from 1 (*all the time*) to 5 (*never*). Higher scores in Ability indicate fewer (self-reported) memory problems in daily situations. The factor Strategy has 19 items related to the use of different memory strategies (e.g., ‘how often do you create a story to link together information you want to remember?’) and respondents indicate the frequency with which each strategy was used over the last 2 weeks using a scale that ranges from 1 (*never*) to 5 (*all the time*). Higher scores in Strategy indicate a more frequent use of memory strategies.

Squire Subjective Memory Questionnaire (SSMQ; Van Bergen, Brands, & Jelicic, 2010; Squire, Wetzel, & Slater, 1979). The SSMQ is an instrument that assesses subjective memory functioning and consists of one single factor reflecting people’s beliefs about their own memory functioning ($\alpha = 0.93$). The instrument includes 18 items (e.g., ‘My ability to reach back in my memory and recall what happened a few minutes ago is’), rated on nine-point scales that range from -4 (*worse than ever before*) to 4 (*better than ever before*). Higher scores in SSMQ indicate a higher self-perceived memory functioning.

Facial recognition and general memory skill assessment (Olsson & Juslin, 1999). Developed by Olsson and Juslin (1999), this instrument includes two items that assess self-reported general memory skill (e.g., ‘give an estimate of your general memory ability, compared to other people’s general memory ability’) and two items that assess self-reported facial recognition skill (e.g., ‘give an estimate of your ability to remember faces as compared to other people’s ability to remember faces’). Participants indicate their ability in comparison to the normal population on a 11-point scale that ranges from -5 (*much worse*) to 5 (*much better*).

Mock crime stimulus. The stimulus event for the eyewitness free recall task was a short film (2:30min) depicting a theft. In the film, two perpetrators (a man and a woman) follow a young man into his house. One of the perpetrators pretends to be lost and asks the victim for directions, and while the victim is distracted the other perpetrator steals his laptop, phone and keys.

Objective Memory Tests

Free-recall test. In the free-recall condition participants completed a second free-recall test for a stimulus which was not associated with the eyewitness paradigm. In this condition, participants watched a short film depicting a cleaning routine in a house (2.5min), then completed an unrelated filler task (5min) before completing a free-recall test concerning this cleaning stimulus film (see Appendix 5).

Cued-recall test. In the cued-recall condition, participants watched the same film of the free-recall test condition depicting a cleaning routine in a house and then completed an unrelated filler task (5min). Next participants completed 17 cued-recall questions about the video (e.g., what did the woman do in the TV room? See Appendix 1). Each question was followed by a confidence scale ranging from 0% (*'not at all certain'*) to 100% (*'totally certain'*).

Face recognition test. Forty-five adult male faces with no unusual identifying features were selected from a database of faces (Thomaz & Giraldi, 2010). Faces were standardized in size, resolution, and background colour. During the training phase, 30 faces were displayed to participants in random order on a computer screen. Each face was presented for three seconds, with a three second inter-stimulus interval. After completing a 5min filler task participants took part in the testing phase, in which a second set of 30 faces was presented, including 15 faces from the training phase and 15 new faces. Each face was shown individually participants were instructed to indicate whether or not the face had been

seen before and give a confidence judgement using a scale that goes from 0% (*'not at all certain'*) to 100% (*'totally certain'*). Participants had unlimited time to make their decision and proceed to the next face.

General knowledge memory test. A pool of 38 general knowledge questions (e.g., 'What country is known as the Land of Rising Sun?') was generated and pilot tested. Eight questions were considered too easy or too difficult and were removed from the final pool. Participants in the General Knowledge condition were asked to answer each one of the 30 questions (see Appendix 2) and rate their confidence in a scale that ranges from 0% (*'not at all certain'*) to 100% (*'totally certain'*).

Procedure

First, using the online platform Qualtrics, participants completed the full set of metamemory measures (MMQ, SSMQ, and Facial recognition and general memory skill assessment, Olsson & Juslin, 1999). Twenty-four hours after completing the metamemory measures, participants took part in a lab session. In this session, participants first watched the mock crime stimulus film and then completed an unrelated filler task (5min). Immediately after participants received the free recall test about the mock crime. The instructions were as follows: "In the space provided, report all details that you can remember about the video, including the sequence of actions and events, and the people that were involved. If you recall information or specific details out of the order in which they happened, report these details as they come to mind (i.e., do not leave out any details.) Do not guess about details that you cannot remember. Feel free to use full sentences or bullet points – but please make sure your report is as complete and accurate as possible". When participants finished the task, the researcher read through the report and marked each detail for which participants should now report a confidence judgement (e.g., A caucasian (1) man (1) in a burgundy (1) hoody (1)

stole (1) the bike (1) = 6 details). Participants were then asked to read through the statement and indicate their degree of confidence for each detail, using a scale that ranged from 0% ('*not at all certain*') and then 10, 20, 30, ... to 100% ('*totally certain*'). After completing their confidence assessment, participants took part in the objective memory test corresponding to their randomly allocated experimental condition (i.e. free recall, cued recall, face recognition or general knowledge tests). Ten per cent of all eyewitness recall transcripts were coded independently by two raters and Intraclass Correlation Coefficients were calculated for number of details = .97 (95% CI = .93, .99), and correctness = .96 (95% CI = .92, .98)

Results

Our analysis focused on four dependent measures in the eyewitness free recall task: accuracy, completeness, confidence and over/underconfidence. Accuracy was defined as the proportion of accurate responses reported, completeness as the total amount of accurate details, and confidence as the mean of all confidence statements provided for each detail. In calibration research, over/underconfidence can be computed as a statistic that relates to how well calibrated participants are in their confidence-accuracy relationship, ranging from -1 (very underconfident) to 1 (very overconfident; see Brewer & Wells, 2006 for calculation details). Similar scores were computed for all objective memory tests (i.e., free recall, cued recall, general knowledge and face recognition). Scores on the objective memory tests served as predictors of eyewitness free recall performance, instead of being treated as dependent variables (see Table 6.1).

Table 6.1

Means and standard deviations of the main memory performance variables

	Proportion Accuracy <i>M (SD)</i>	Completeness <i>M (SD)</i>	Confidence <i>M (SD)</i>	Over/underconfidence <i>M (SD)</i>
Eyewitness free recall (<i>n</i> = 208)	0.97 (0.04)	40.7 (11.4)	0.92 (0.58)	-0.04 (0.06)
Free recall (<i>n</i> = 55)	0.95 (0.05)	30.2 (13.2)	0.93 (0.52)	-0.02 (0.06)
Cued recall (<i>n</i> = 50)	0.69 (0.11)	-	0.64 (0.14)	-0.05 (0.13)
Face recognition (<i>n</i> = 50)	0.45 (0.08)	-	0.43 (0.92)	-0.01 (0.15)
General Knowledge (<i>n</i> = 55)	0.41 (0.15)	-	0.39 (0.18)	-0.02 (0.14)

Metamemory Scale Scores And Eyewitness Free Recall Performance

First, we fitted a multivariate regression model including the scale scores on the metamemory assessments as predictors of free recall accuracy, completeness, confidence and over/underconfidence. One outlier in the completeness variable was highly influential in the test parameters (Cook's Distance = 0.13) and was capped to the upper limit to avoid biased results. QQ-plots revealed that the errors distribution in the accuracy and confidence models were not normally distributed. Therefore, we fitted these models using gamma distributions in order to obtain more robust estimates (Manning, Basu, & Mullahy, 2005). None of the metamemory scales were predictive of accuracy, completeness, confidence or over/underconfidence in the eyewitness free recall test (smallest *p* value = .17). Results from four different multivariate tests (Pillai, Wilks, Hotelling-Lawley, and Roy) suggest that the coefficients for MMQ-Contentment, MMQ-Strategy, MMQ-Ability, SSMQ, and the cognitive self-assessment by Olsson and Juslin (1999) do not seem to be statistically different from 0 for any of the dependant measures (see Table S6.1 in Appendix 6). Bayesian multiple regression models were fitted to further examine evidence for the null model, using a

standard Jeffreys–Zellner–Siow prior (see Table 6.2). Results revealed that marginal inclusion probabilities for all coefficients in the accuracy, completeness, confidence and over/underconfidence models were negligible (highest inclusion probability = .07; Kruschke, 2015).

Table 6.2

Bayesian Multiple linear regression models including metamemory scales as predictors of eyewitness free recall performance

	95% Credible Interval				
Coefficient	<i>M</i>	<i>SD</i>	BF _{inclusion}	Lower	Upper
Free recall accuracy					
MMQ-Contentment	-1.033e -4	9.148e -4	0.049	0.000	0.000
MMQ-Ability	7.298e -5	9.946e -4	0.045	0.000	0.000
MMQ-Strategies	-1.113e -4	9.914e -4	0.049	0.000	0.000
SSMQ	-5.650e -7	4.521e -4	0.042	0.000	0.000
General Memory Ability	4.057e -6	3.028e -4	0.042	0.000	0.000
Face Memory Ability	-1.961e -5	2.760e -4	0.044	0.000	0.000
Free recall accuracy					
MMQ-Contentment	-0.018	0.273	0.047	0.000	0.000
MMQ-Ability	-0.043	0.383	0.052	-0.110	0.000
MMQ-Strategies	0.071	0.423	0.065	0.000	0.611
SSMQ	0.016	0.177	0.049	0.000	0.000
General Memory Ability	-0.002	0.105	0.045	0.000	0.000
Face Memory Ability	0.012	0.105	0.052	-0.085	0.000
Free recall confidence					
MMQ-Contentment	-0.003	0.033	0.186	-0.100	0.068
MMQ-Ability	0.055	0.088	0.596	0.000	0.261
MMQ-Strategies	-0.007	0.035	0.193	-0.135	0.021
SSMQ	0.012	0.032	0.302	-0.011	0.090
General Memory Ability	0.012	0.026	0.418	-0.007	0.090
Face Memory Ability	0.018	0.027	0.678	-0.002	0.077
Free recall over/underconfidence					
MMQ-Contentment	2.426e -4	0.002	0.064	-8.178e -4	3.309e -4
MMQ-Ability	-0.002	0.005	0.132	-0.017	0.000
MMQ-Strategies	1.985e -4	0.002	0.061	-8.374e -5	0.002
SSMQ	1.333e -4	0.001	0.059	-3.329e -4	5.241e -4
General Memory Ability	5.799e -5	8.374e -4	0.057	-1.340e -4	5.694e -4
Face Memory Ability	-4.606e -5	6.504e -4	0.056	-2.721e -4	2.763e -4

Objective Memory Tests And Eyewitness Free Recall Performance

We used simple regressions to test the relationship between eyewitness free recall performance and performance on the four different memory tests (see Table 6.3). Eyewitness free recall accuracy was predicted by cued recall accuracy ($\beta = 0.38$), but not by accuracy in the other memory tests. Confidence in the eyewitness free recall was predicted by confidence expressed in the free recall ($\beta = 0.75$), cued recall ($\beta = 0.54$) and general knowledge memory tests ($\beta = 0.30$). A similar pattern of results was observed for over/underconfidence in the free recall test, which was predicted by over/underconfidence in the free recall ($\beta = 0.63$), cued

Table 6.3

Regression models using performance in the different memory tests as predictors of eyewitness free recall performance

	<i>B</i> (<i>SE</i>)	<i>B</i> CI	β	<i>t</i>	<i>p</i>
Eyewitness recall completeness					
Free recall completeness	0.57 (0.07)	[0.41, 0.72]	0.71	7.44	<.001
Eyewitness recall accuracy					
Free recall accuracy	0.12 (0.10)	[-0.08, 0.34]	0.16	1.23	.22
Cued recall accuracy	0.14 (0.05)	[0.04, 0.25]	0.38	2.92	<.01
General Knowledge accuracy	-0.04 (0.02)	[-0.07, 0.01]	-0.21	-1.59	.12
Face recognition accuracy	-0.04 (0.04)	[-.12, 0.05]	-0.12	-0.84	.40
Eyewitness recall confidence					
Free recall confidence	0.83 (0.09)	[0.63, 1.03]	0.75	8.34	<.001
Cued recall confidence	0.25 (0.05)	[0.13, 0.36]	0.54	4.49	<.001
General knowledge confidence	0.09 (0.04)	[0.01, 0.17]	0.30	2.27	.02
Face recognition confidence	-0.05 (0.08)	[-0.23, 0.12]	-0.08	-0.60	.54
Eyewitness recall OU					
Free recall OU	0.70 (0.11)	[0.46, 0.94]	0.63	5.84	<.001
Cued recall OU	0.29 (0.06)	[0.17, 0.42]	0.56	4.71	<.001
General knowledge OU	0.24 (0.06)	[0.12, 0.36]	0.49	4.03	<.001
Face recognition OU	0.08 (0.06)	[-0.04, 0.21]	0.17	1.26	.21

recall ($\beta = 0.56$) and general knowledge ($\beta = 0.49$) tests. Performance in the face recognition test did not relate to any of the performance variables in the eyewitness free recall.

Metamemory And Confidence-Accuracy Calibration Analysis

Calibration analysis were conducted to test the relation between the metamemory scale scores and eyewitness confidence-accuracy relationship in the free recall test. We first produced calibration curves for all the memory tests (see Figure 6.1). The diagonal line represents perfect calibration, such that each level of confidence is equivalent to the level of accuracy for decisions made with that level of confidence. Observations above this line indicate underconfidence, and observations below this line indicate overconfidence. Three calibration statistics were computed: calibration index, over/underconfidence and resolution. Calibration (C) represents how far a given calibration curve is from a perfect calibration. It ranges from 0 (perfect calibration) to 1, and lower values indicate better calibration. Over/underconfidence (O/U) indicates if a curve strays more above or below the perfect calibration line, with values ranging from -1 (very underconfident) to 1 (very overconfident). The Adjusted Normalized Resolution Index ($ANRI$) represents how well confidence discriminates accurate from inaccurate identifications, with higher values indicating better discrimination (see Brewer and Wells (2006). Following Palmer, Brewer, Weber and Nagesh (2013), we used a jackknife procedure to compute standard errors for each calibration statistic, which were then converted to 95% inferential confidence intervals (Tryon, 2008). If the confidence intervals do not overlap, that indicates a significant difference (see Table 6.4). The C statistic pointed to a reasonably strong calibration for the eyewitness free recall test ($C = .02$). However, examining the calibration curves it can be observed that most information disclosed was accurate, even those reported with low levels of confidence, so that confidence and accuracy in the eyewitness free recall did not co-vary systematically (see Figure 6.1). Performance in the face recognition was the most distant from perfect calibration, presenting underconfidence for lower levels of confidence and overconfidence for higher levels of confidence.

Table 6.4

Calibration statistics for each memory test with inferential confidence intervals (ICI)

	<i>C</i> [ICI]	<i>OU</i> [ICI]	<i>ANRI</i> [ICI]
Eyewitness recall	0.02 [0.01,0.02]	-0.05 [-0.05, -0.04]	0.04 [0.02, 0.06]
Free recall	0.01 [0.01,0.01]	-0.02 [-0.03, -0.01]	0.09 [0.02, 0.15]
Cued recall	0.02 [0.01, 0.03]	-0.05 [-0.08, -0.02]	0.25 [0.19, 0.31]
General Knowledge	0.01 [0.01, 0.02]	-0.01 [-0.03, 0.01]	0.35 [0.30, 0.40]
Face recognition	0.07 [0.06, 0.09]	-0.03 [-0.06, 0.00]	0.01[-0.01,0.01]

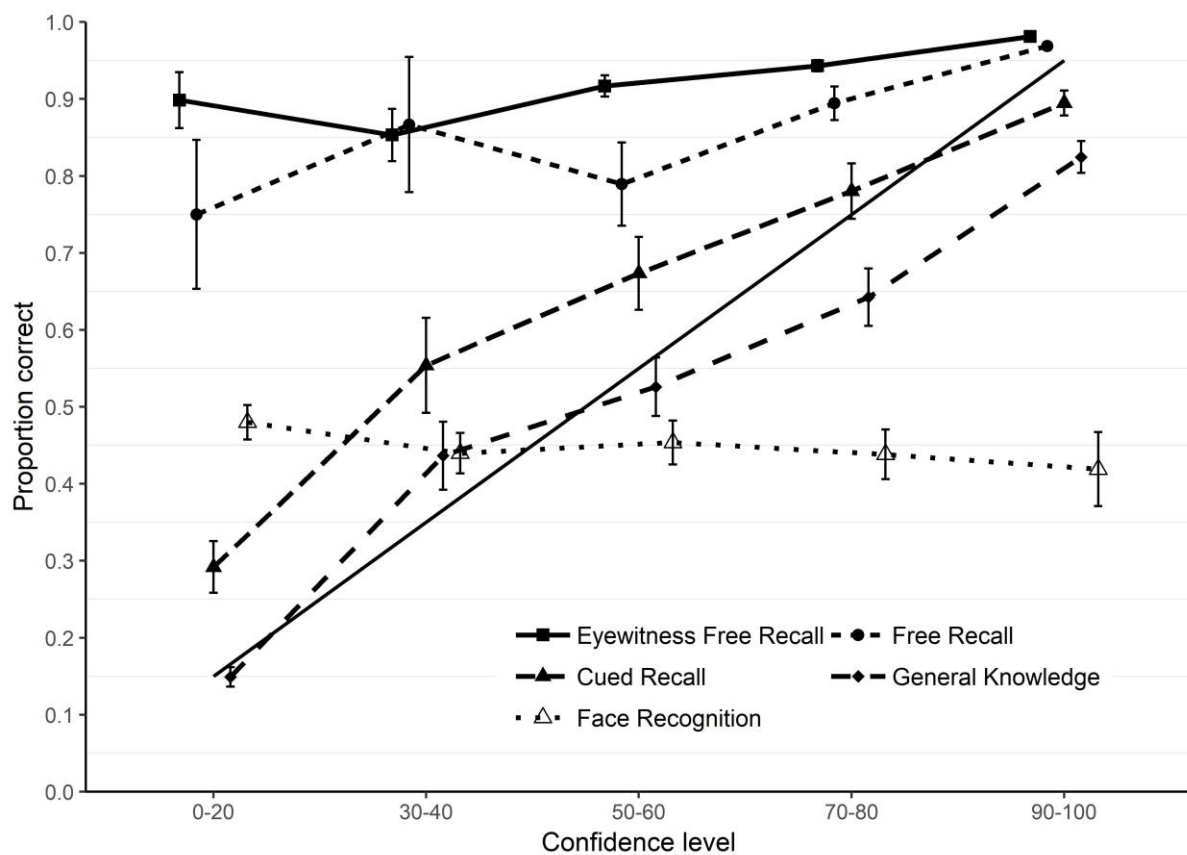


Figure 6.1. Calibration curves of eyewitness free recall and all memory tests. Error bars are 95% confidence intervals.

Next, we compared calibration statistics between high and low scorers on each of the metamemory measures. Following a procedure similar to the one used by Olsson and Juslin (1999), individuals above the 66th percentile were selected as high scorers and individuals below the 33th percentile were selected as low scorers (see Figure 6.2). Inspection of the confidence intervals suggest that high scorers in the MMQ-Contentment, MMQ-Ability, SSMQ, and General Memory Skill were less underconfident and slightly better calibrated than low scorers in those components. There were no observable differences between the resolution scores of high and low metamemory scorers.

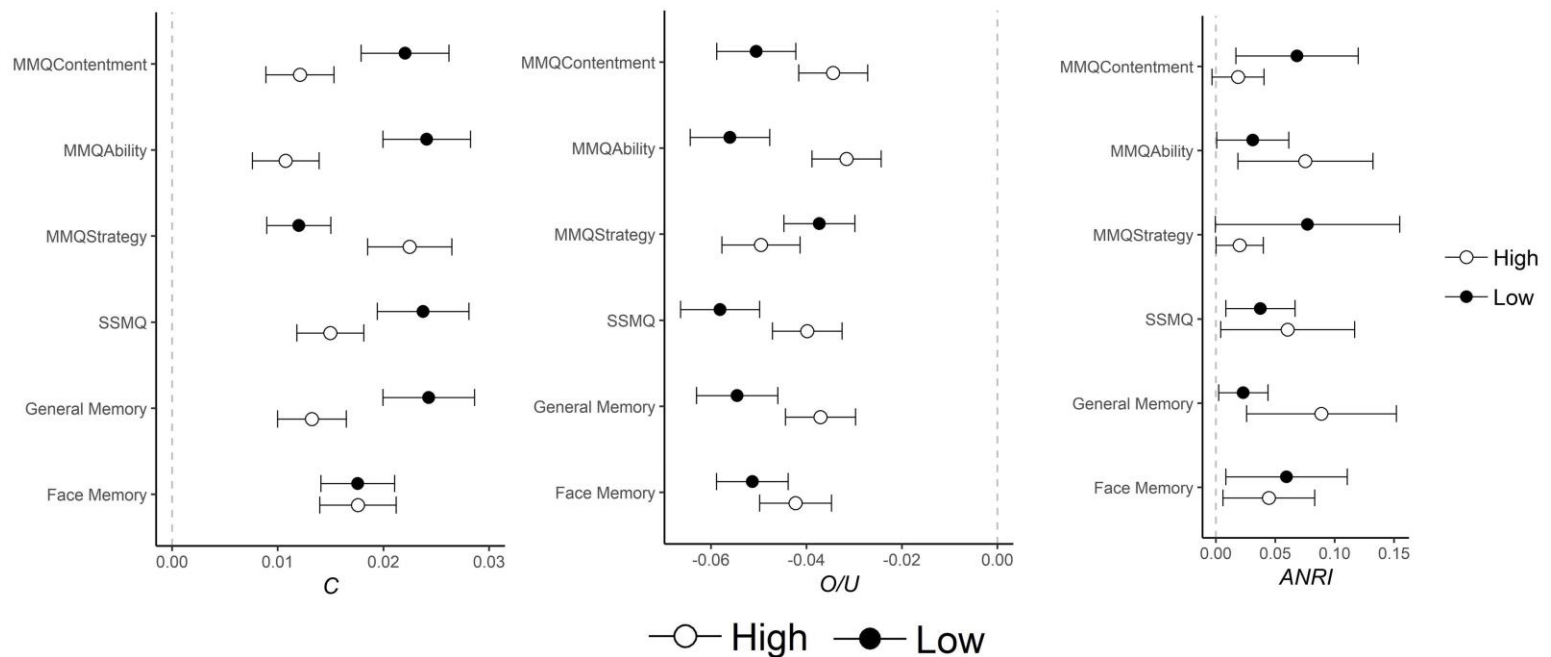


Figure 6.2. Inferential confidence intervals of calibration statistics for high and low scorers in each metamemory measure.

Discussion

Recently there has been great interest in the predictive utility of subjective and objective memory assessments for eyewitness performance (e.g., Baldassari, Kantner, & Lindsay, 2019; Russ, Sauerland, Lee, & Bindemann, 2018). In the current study testing the

use of metamemory measures and memory tests as predictors of eyewitness free recall performance, we contribute three key findings: (1) Contrary to some of our hypotheses, the metamemory scales examined had no relation with accuracy, completeness, confidence or over/underconfidence in eyewitness free recall; (2) Individuals with high metamemory scores presented a slightly stronger confidence-accuracy calibration in eyewitness free recall; and, (3) Eyewitness free recall confidence and over/underconfidence was predicted by performance in objective memory tests. These findings extend our understanding of how eyewitness performance relates to subjective self-assessments of memory ability (Bornstein & Zickafoose, 1999; Evans & Fisher, 2011) and objective memory performance in different tasks (Bindemann et al., 2012; Morgan et al., 2007).

We predicted that self-assessments of memory capacity and other metamemory components would be related to eyewitness accuracy, completeness, confidence and over/underconfidence. Previous studies have found positive associations between eyewitness free recall performance and individual differences such as working memory capacity, intelligence, and temperament (Chae & Ceci, 2005; Jaschinski & Wentura, 2002; Zhu et al., 2010). In the current study, we adopted diverse measures assessing different components of metamemory, such as memory ability, endorsement of memory strategies, and memory development over time. Surprisingly, there is no evidence that the metamemory scales used related to eyewitness free recall performance, suggesting that intrinsic cues of general memory capacity may have no relation with the completeness or quality of eyewitness memory reports. The metamemory scales are also unrelated to the confidence or over/underconfidence in the eyewitness free recall, suggesting that confidence expressed for freely recalled information may originate mostly from memory trace strength, rather than self-credibility cues. Some theoretical frameworks propose that individuals freely recalling a witnessed event monitor their memories and control what they report in order to achieve an

acceptable accuracy (Evans & Fisher, 2011; Koriat & Goldsmith, 1996). Koriat, Nussinson, Bless, and Shaked (2008) argue that monitoring process may be based on domain-specific beliefs (e.g., “I do not have a very good memory”) or on the learning experience (i.e., the experience of processing and remembering the learning material). Our findings demonstrate that learning experience may be more influential than domain-specific beliefs when eyewitnesses are monitoring which and how much information to disclose during free recall.

Apart from the finding that metamemory measures are not predictive of free recall over/underconfidence on the individual level, we observed that individuals who score high on some metamemory components have a stronger confidence-accuracy relationship than low scorers in the group level. Specifically, individuals who claim to have higher memory contentment, self-rated memory ability, and memory development over time are slightly less underconfident than those individuals with lower scores in the same components. So, at least at the group level, individuals with higher levels of contentment with their own general memory ability may present confidence statements that better reflect their probable accuracy in an eyewitness free recall task. In other words, individuals who are not content with their own memories may present lower confidence statements even if they are probably accurate. That finding is consistent with previous results by Olsson and Juslin (1999), who observed that individuals who rated themselves as good face recognizers demonstrated a more diagnostic confidence-accuracy relation. It is important to note, however, that in the current research, this effect was mostly associated to lower level of confidence, given that responses with high levels of confidence were almost invariantly correct. That is, high levels of confidence in the free recall tasks were almost always associated with correct information. Furthermore, the associations between those metamemory measures (i.e., memory contentment, self-rated memory ability, and memory development) and eyewitness over/underconfidence were only observed on the group level (i.e., low raters vs high raters),

and no direct relation was observed when testing the same measures as predictors over/underconfidence.

In contrast to subjective self-ratings of memory ability, stable individual differences in objective memory performance may be a better indicator of eyewitness free recall performance. In fact, in some studies it was found that performance in face recognition tests is somewhat predictive of eyewitness lineup performance (e.g., Baldassari et al., 2019; Russ et al., 2018). In the current study, we predicted that performance in different memory tests would be related to performance in an eyewitness free recall task. The results show that completeness in an eyewitness free recall was closely related to the completeness of an unrelated free recall. This finding seems to indicate some stability in the reporting of accurate information across different testing situations, which may be explained by individual differences related to attentional and cognitive resources, engagement with the task, or a combination of both (Aslan & Bäuml, 2011; Melby-Lervåg, Redick, & Hulme, 2016).

We also observed that confidence and over/underconfidence in the free recall, cued recall and general knowledge tests were predictors of confidence and over/underconfidence in the eyewitness free recall test. This finding corroborates previous findings showing that individuals express confidence in a somewhat stable manner across different testing conditions, including eyewitness memory domain (Bornstein & Zickafoose, 1999; Jonsson & Allwood, 2003; Mengelkamp & Bannert, 2010). Jonsson and Allwood (2003), for example, found some individual stability in confidence judgements for word knowledge and logical/spatial ability tasks, while Bornstein and Zickafoose (1999) found that overconfidence, calibration, and resolution in a general knowledge and cued recall test were positively correlated. We expand on those findings by presenting evidence of individual stability not only for confidence, but also for over/underconfidence in eyewitness free recall and unrelated tests (i.e., free recall, cued recall and general knowledge). Interestingly, this

association was stronger for memory tests more closely related to the eyewitness free recall task (e.g., free recall, followed by cued recall and general knowledge). Complementarily, there was no stability between the eyewitness free recall and face recognition task, a finding in line with models outlining independent systems for face, episodic and semantic memory (Yovel & Kanwisher, 2004). Such consistency in individual's confidence and overconfidence for similar memory domains may be explained by stable metacognitive cues (e.g., previous experience in that memory domain (Koriat et al., 2008). Taken together, our results support frameworks that propose domain-specific memory self-efficacy, defined as an individual's appraisal of his or her usual ability in a given memory domain (Beaudoin & Desrichard, 2011; Hertzog & Dixon, 1994).

Eyewitness statements are commonly sought to aid investigations or to obtain incriminatory or exculpatory evidence. We found no indication for a relation between eyewitness free recall performance and metamemory self-assessments, including self-rated memory ability, endorsement of memory strategies and memory development over time. This finding may contribute to two theoretical predictions requiring further research: (1) individuals may have limited insight on their own *general* memory ability (Beaudoin & Desrichard, 2011; Perfect, 2004); or (2) individuals may have accurate insight on their *general* memory ability, but such intrinsic cues have little influence on the disclosure of witnessed events (Evans & Fisher, 2011; Koriat & Goldsmith, 1996). From a practical perspective, it is problematic to question the credibility of eyewitness reports based on general self-assessments of memory capacity. As argued by Evans and Fisher (2011), contrary to common belief eyewitnesses who say 'I don't know' more often may be more accurate than other witnesses, given that they are better monitoring their reporting to provide accurate information. Additionally, we observed that on the group level individuals who distrust their own memories have a weaker confidence-accuracy relationship, presenting

lower confidence statements even if they are likely accurate. If replicated, this finding could offer a basis for metamemory assessments to better discriminate overconfident and underconfident eyewitnesses. Finally, we found some stability between eyewitness free recall performance and performance in related memory tasks (i.e., unrelated free recall, cued recall and general knowledge). This result lends some support to the concept of an objective test designed to estimate eyewitness free recall performance, an approach that has been tested for eyewitness identification settings (Baldassari et al., 2019; Bindemann et al., 2012; Morgan et al., 2007).

There are a number of limitations that should be taken into consideration. First, the mock crime video was seen in the same session as the eyewitness free recall test. This procedure was adopted to guarantee the feasibility of the study, but in more naturalistic contexts eyewitnesses commonly report what they have seen after a longer period. Importantly, previous studies have found differential effects of self-capacity measures and stability in memory tests that were dependant on task difficulty (Howie & Roebbers, 2007; Mengelkamp & Bannert, 2010; Stankov, 2000). It may be the case, for example, that the relation between objective memory tests and eyewitness free recall performance is weaker if there is significant memory decay for information about the crime. Furthermore, we observed high accuracy and high confidence in the eyewitness free recall reports, which may have been due to the nature of the task or because the interval between encoding and retrieval was short making the task easy. We highlight the need for research that investigates associations between measures of self-efficacy, performance in objective memory tests, and eyewitness free recall performance under varying levels of difficulty (e.g., longer retention intervals; Sauer et al., 2010). A second limitation is that the metamemory assessment always occurred 24 hours before the eyewitness paradigm. In naturalistic contexts, such metamemory assessment would realistically occur after an eyewitness account was obtained. In

determining our procedure, we reasoned that exposure to the eyewitness paradigm and other memory tests before the completion of the metamemory measures would have affected witnesses' self-assessments of their memory abilities to a greater degree than completing the assessments would affect eyewitness performance. While this was perhaps appropriate for an initial test of relationships between the tasks, future studies should examine the use of metamemory measures after the eyewitness task.

Conclusions

In sum, our results allow for three main inferences. First, metamemory assessments appear to have little value in estimating individual performance in eyewitness free recall settings. With reference to utility in the applied context, this finding suggests that dismissing or questioning the credibility of eyewitness reports based on self-assessments of memory capacity may be unwarranted. Second, at the group level individuals with higher self-ratings of memory capacity had a slightly stronger confidence-accuracy calibration. This is initial evidence that free recall confidence is a better predictor of accuracy among individuals that are not very doubtful about their own memory performance. Third, we find stability in confidence and over/underconfidence measures across eyewitness free recall and other memory tests of similar domain. If this relationship is replicated in future research it may indicate that individual differences or intrinsic metamemory cues (e.g., experience with memory issues) partly explain levels of realism for confidence judgements in memory tasks.

Data Availability Statement

All datasets and R analysis scripts for this study can be found at:

https://osf.io/cywtk/?view_only=aa652cf2bc7f4fdd88a6b96053192bb7

CHAPTER 7

General Discussion

Summary of the Findings

The overarching aim of the current thesis was to examine the relationship between metamemory self-assessments and different aspects of eyewitness testimony performance. Of particular interest was whether metamemory assessments might be used to distinguish accurate from inaccurate eyewitness evidence. Eyewitness evidence plays a major role in the criminal justice system, but inaccurate witnesses can impair investigative processes and even contribute to miscarriages of justice. Theories of metamemory provide a framework for understanding how individuals formulate an informative account of a past event and regulate the reporting of information (Koriat & Goldsmith, 1996; Koriat, Goldsmith, & Pansky, 2000). However, to date, the intersection between metamemory research and eyewitness testimony has not been systematically explored. To address this gap, all experiments in the current thesis focused on how eyewitness metamemory traits relate to accuracy, confidence, and confidence-accuracy calibration in identification tasks and free recall paradigms.

In the first two experiments (Experiments 1 and 2, **Chapter 2**), we examined the predictive value of metamemory assessments and objective memory tests as estimators of eyewitness identification accuracy, confidence, and confidence-accuracy calibration. Results suggested that found that general memory contentment (MMQ-Contentment, e.g., “I am generally pleased with my memory ability”) was positively related to eyewitness identification accuracy among choosers, with no relation for nonchoosers. Additionally, it was observed that confidence expressed in identification tasks was stable across memory tasks of similar domain (i.e., face recognition). That is, confidence expressed in different memory tests was related to the confidence expressed in eyewitness identifications. In Experiment 3 (**Chapter 3**), we sought to expand validity of our metamemory assessment by developing the Eyewitness Metamemory Scale, a psychometric instrument measuring self-rated ability and endorsement of strategies for face recognition and person identification. In

Experiment 4 (**Chapter 4**), we examined eyewitness-specific and general metamemory assessments as predictors of eyewitness identification performance for biased and unbiased lineups. We found that EMS-Discontentment (e.g., “Sometimes I have trouble recognizing a person that I know relatively well”) was predictive of identification accuracy for both biased and unbiased lineups, and individuals who scored lower in eyewitness metamemory factors display a stronger confidence-accuracy calibration than those who had higher scores. In Experiment 5 (**Chapter 5**), we aimed to extend these findings by adopting a repeated-trial eyewitness identification paradigm, in order to compute confidence-accuracy calibration scores for each participant in the experiment. In this study, we found that EMS-Contentment (e.g., “My ability to remember faces is much better than other people’s ability to remember faces”) was predictive of identification accuracy, confidence, and confidence-accuracy calibration. Finally, in Experiment 6 (**Chapter 6**) we examined the relationship between metamemory assessments and the reporting of information in eyewitness free recall. Results from this experiment showed that individuals with higher self-rated memory capacity had a slightly stronger confidence-accuracy relation in free recall, although subjective assessments of memory capacity were not related to the amount or accuracy of reported information. Taken together, these experiments provide some converging evidence that metamemory traits are associated with eyewitness memory performance in terms of accuracy, confidence, and confidence-accuracy relation.

We now briefly summarize and integrate the main findings of each chapter in terms of the overarching goal of this thesis: to elucidate the relationship between metamemory and eyewitness testimony performance. Several aspects of eyewitness performance were examined across the different experiments in this thesis. We first discuss the relationship between metamemory assessments and eyewitness memory accuracy. Next, we discuss the relationship between metamemory traits and expressions of confidence, linking our results to

current theories concerning how confidence judgements are formed and reported. We then focus on the relationship between metamemory traits and the confidence-accuracy relationship in eyewitness settings. We also consider the relation between metamemory traits and reporting of information in eyewitness free recall. Finally, we discuss the results on how different objective memory tests (i.e., general knowledge, cued-recall, free-recall and face recognition) relate to eyewitness performance in identification and free recall tasks. The limitations of the current research and emerging findings will then be considered. Finally, we elaborate on future directions and the practical implications for this line of research.

Metamemory and Identification Accuracy

One of the main goals in the current thesis was to determine whether metamemory self-assessments might be used to effectively distinguish accurate from inaccurate eyewitness identifications. Although witness evidence can help solve crimes, inaccurate identifications can jeopardize investigations and place innocent suspects at risk of prosecution. In what we believe to be among the first set of experiments investigating the role of metamemory in identification performance, we found that self-ratings of memory ability are useful estimators of eyewitness identification accuracy (Experiments 1, 2, 4, and 5). That is, in our two initial experiments (Experiments 1 and 2), we found that an individual's memory contentment with their general memory ability (MMQ-Contentment) was predictive of identification accuracy among choosers. In later experiments we expanded on this finding and observed that self-ratings of memory contentment that are specific to face recognition and person identification are better predictors of identification accuracy compared to general metamemory assessments (Experiments 3 and 4).

These findings are consistent with research demonstrating a linkage between memory self-efficacy (MSE) and objective memory performance (Seeman, Rodin, & Albert, 1993;

Seeman, McAvay, Merrill, Albert, & Rodin, 1996; Valentijn et al., 2006). Olsson and Juslin (1999), for example, found that individuals who considered themselves to be better face recognizers were more accurate in identification tasks. Other research has found that individuals have an accurate, albeit limited insight into their own ability to recognize faces (Bindemann, Attard, & Johnston, 2014; Bobak, Mileva, & Hancock, 2018). The data from the current programme of work suggest that memory self-efficacy (MSE) assessed by metamemory tools may be reliable indicators of identification performance, especially when the assessed MSE is specific to face recognition and person identification.

The pattern of results observed in Experiments 1 to 5 are consistent with a domain-specific explanation of the relation between memory function for a specific memory domain (e.g., face recognition) and an individuals' self-perceived capacity to perform a memory task on that same domain. A domain-specific MSE can be defined as an individual's appraisal of his or her usual ability to memorize and retrieve one particular kind of information (e.g., "I am good at remembering faces"; Hertzog & Dixon, 1994). The Contentment factor in the Eyewitness Metamemory Scale, for example, is used to assess individuals' perceptions of their face recognition and person identification ability. In contrast, global MSE refers to self-assessments of memory capacity across many different memory domains and tasks (e.g., remembering proper nouns, telephone numbers, appointments), without reference to any specific context (Berry, 1999; Cavanaugh & Green, 1990). The Contentment factor of the Multifactorial Metamemory Questionnaire is one example of a factor-analytical scale used to assess individuals' general perceptions of their memory abilities. It requires respondents to indicate their level of agreement (from strongly disagree to strongly agree) with 18 judgments such as "I am generally pleased with my memory ability" or "I have confidence in my ability to remember things".

According to MSE theory, both global and domain-specific memory self-efficacy are aggregates of individuals' appraisals of their usual memory abilities over diverse memory tasks (Hertzog, Dixon, & Hultsch, 1990). The same theory predicts that perceived ability is likely to vary from one memory task to another, depending on individual differences and on the particular features of each task. Because perceived ability varies from one memory domain to another, global MSE scores tend to be less predictive than domain MSE scores of performance on a given task (Beaudoin & Desrichard, 2011; Crook & Larrabee, 1990; Christopher Hertzog, 2002). Our findings across Experiments 1 to 5 corroborate this prediction, showing evidence that general MSE has little predictive value for eyewitness identification performance compared to domain MSE specific to face recognition and person identification, as measured by the Eyewitness Metamemory Scale. Specifically, we only observed a significant relation between global MSE (i.e., MMQ-Contentment) and eyewitness identification accuracy in Experiments 1 and 2, and this finding was not replicated in Experiments 4 and 5. Instead, in those later experiments we found evidence that discontentment (Experiment 4) and contentment (Experiment 5) with face recognition and person identification ability was predictive of eyewitness identification accuracy.

This pattern of results reveals important theoretical and practical implications concerning the relation between MSE and performance in memory-demanding tasks in forensic settings. From a theoretical perspective, these findings highlight the importance of clearly defining separate systems of MSE taking into account specific memory domains. Theoretical models of MSE that account for specific memory domains allow for more precise predictions of the relation between MSE and memory performance (Beaudoin & Desrichard, 2011; Christopher Hertzog, 2002). From a practical perspective, our findings suggest that MSE assessments that are specific to face recognition and person identification are useful indicators of eyewitness identification accuracy. Importantly, in Experiment 4 we find that

self-ratings of face recognition ability are indicative of identification accuracy for both biased and unbiased lineups. This finding reinforces the potential diagnostic value of MSE assessments given that other postdictors of identification accuracy (e.g., eyewitness confidence and decision time) tend to have lower diagnostic value in situations involving biased lineups (Charman, Wells, & Joy, 2011; Key et al., 2017). Although the current findings require further replication before any potential application in the field is considered, these observations highlight the potential applied value of domain-specific MSE assessments to distinguishing accurate from inaccurate identifications.

Metamemory and Identification Confidence-Accuracy Relationship

The level of confidence expressed by an eyewitness affects how investigators and triers of fact perceive the likely accuracy of the eyewitness account or identification. Eyewitnesses who confidently indicate a suspect as the perpetrator of a crime are perceived as more reliable than those that appear uncertain about what they remember (Brewer & Burke, 2002; Wells, Olson, & Charman, 2002). In many circumstances, confidence is indeed a reliable indicator of witness accuracy, but confidence can also be contaminated by several external factors, undermining its diagnostic value (Palmer, Brewer, Weber, & Nagesh, 2013; Sauer, Brewer, Zweck, & Weber, 2010; Wixted & Wells, 2017). In the experiments conducted for this thesis, we found that confidence judgements expressed in identification tasks are also partially influenced by internal cues related to self-assessments of memory efficacy. More specifically, we observed that assessments of MSE that are specific to face recognition are better predictors of identification confidence compared to general assessments of MSE.

In Experiments 1 and 2 (**Chapter 2**), we found no evidence that general assessments of MSE were related to confidence expressed by witnesses in identification tasks. However,

in Experiment 5 (**Chapter 5**), we found that MSE that is specific to eyewitness identification settings (as measured by the Contentment and Discontentment factors of the Eyewitness Metamemory Scale developed in Experiment 3, **Chapter 3**) was predictive of witnesses' identification confidence. That is, witnesses with higher MSE for face recognition and person identification tended to be more confident in their identification decisions, while individuals with lower MSE for face and person identification tended to be less confident.

This pattern of results aligns with current theories of metamemory judgments which postulate that memory confidence in a given task is based on both external and internal cues. One influential theoretical framework, for example, predicts that metacognitive judgements are based on cues related to information (e.g., perceived difficulty of the task) and experience (e.g., previous memory performance in similar tasks; Koriat, 2000). In fact, a common assumption across different theories of metamemory judgements is that confidence originates not only from memory vividness or completeness but also from domain knowledge and beliefs about one's own memory efficacy (Dunlosky & Bjork, 2008; Metcalfe, 2009; Metcalfe, Schwartz, & Joaquim, 1993). Various studies have demonstrated a relationship between internal cues such as MSE and expressions of confidence (e.g., Mueller, Tauber, & Dunlosky, 2013), but investigations on the relationship between MSE and confidence judgements in eyewitness identification settings are sparse.

In eyewitness testimony research it has, however, been demonstrated that witnesses' identification confidence can be contaminated by external factors. For example, an eyewitness who receives positive identification feedback after an identification decision tends to be more confident when asked later about the identification decision (Bradfield, Wells, & Olson, 2002; Palmer, Brewer, & Weber, 2010; Wells & Bradfield, 1999). Moreover, eyewitnesses tend to be more overconfident if implausible lineup fillers are used (Charman et al., 2011) or if repeated identifications are conducted (Granhag, 1997; Steblay & Dysart,

2016). In the experiments comprising the current thesis, we present some of the first evidence that confidence judgements in eyewitness identification tasks are not only influenced by external factors, but also by intrinsic cues related to one's own perception of their memory performance.

This finding has important implications for theory and practice because it demonstrates that some witnesses may tend to be overconfident or underconfident depending on their memory self-efficacy. Overconfident witnesses who make false identifications can put innocent suspects at higher risk of prosecution, while underconfident witnesses, despite making correct identifications, may mislead investigators to overlook correct lines of investigation. It has been suggested that confidence is a reliable indicator of identification accuracy, but only if external factors that may contaminate confidence are not present in the case at hand (Wixted, Mickes, & Fisher, 2018; Wixted & Wells, 2017). We expand on this discussion by showing that witnesses confidence statements are not only affected by external factors but also by internal cues related to individual differences. That is, we find that confidence statements in eyewitness identifications are partially based on assessments of memory self-efficacy. A question that arises from this observation is whether assessments of memory self-efficacy may be used to distinguish overconfident from underconfident witnesses during identification tasks.

Olsson and Juslin (1999) found that witnesses who claimed to have good face recognition ability also had better confidence-accuracy calibration than those claiming to have a poor memory for faces. However, that particular piece of research used a two-item scale of self-rated face recognition ability of unknown validity and reliability. In our experiments we aimed to adopt more valid measures of both general memory self-efficacy and face recognition self-efficacy to estimate its relationship with eyewitness's confidence-accuracy relation in identification tasks. Overall, we found little evidence that general

assessments of MSE can distinguish individuals with strong or weak confidence-accuracy relation. Specifically, in Experiments 1 and 2 (**Chapter 2**) only the Strategies factor of the Multifactorial Metamemory Questionnaire (a measure of general metamemory) was related to eyewitness confidence-accuracy relation, so that individuals with higher endorsement of general memory strategies (e.g., “How often do you write down in a notebook things that you want to remember?”) were better calibrated than individuals with lower scores in the same factor. In the other experiments, there was some evidence that eyewitness-specific ratings of MSE may be more strongly related to confidence-accuracy relation in identification tasks. In Experiment 4 (**Chapter 4**), lower scores in the EMS-Contentment factor were predictive of better confidence-accuracy relationship among choosers for both biased and unbiased lineups. However, this same result was not consistent in Experiment 5 (**Chapter 5**), where higher scores in EMS-Contentment were indicative of better confidence-accuracy calibration. This discrepancy may be explained by the fact that analysis in Experiment 4 (**Chapter 4**) was separated by choosers and nonchoosers, while analysis in Experiment 5 (**Chapter 5**) consisted of an aggregate model including performance for both choosers and nonchoosers. This different analytical approach was used because in Experiment 5 participants completed multiple lineup identifications, and in this case separation by choosers and nonchoosers would result in duplicate participants in each group (see **Chapter 5**).

Although the exact relationship between MSE and the confidence-accuracy relationship remains to be determined, the experiments in the current thesis provide initial evidence that individual differences in face memory self-efficacy are related to witnesses confidence-accuracy relation in identification tasks. This contribution adds to an emerging literature on individual differences as postdictors of eyewitness identification performance (Baldassari, Kantner, & Lindsay, 2019; Bindemann, Brown, Koyas, & Russ, 2012; Bobak et al., 2018). Some studies, for example, have found a moderate relationship between objective

memory performance in standardized face recognition tests and eyewitness identification accuracy (Bindemann et al., 2012; Morgan et al., 2007). With the current thesis, we provide a foundation for further investigations regarding the use of MSE assessments to estimate how confident or underconfident an eyewitness is likely to be in an identification task. One advantage of self-assessments of memory capacity over standardized tests of memory performance is that they can be more easily applied in a variety of contexts, given that standardized tests may be more time consuming and not readily available. Further research examining the potential of MSE assessments to distinguish overconfident from underconfident witnesses is specifically warranted given that witness confidence is both malleable and highly influential in court decisions (Brewer & Burke, 2002; Wells et al., 2002). Importantly, in many circumstances it is difficult to estimate how many external factors may have negatively influenced an eyewitness confidence in a given case (e.g., occurrence of positive identification feedback or biased lineups; Charman et al., 2011; Wells & Bradfield, 1999). If MSE assessments can reliably predict an eyewitness susceptibility to overconfidence or underconfidence, it may be used at any point in the evaluation of an eyewitness identification to better decide on the diagnosticity of this evidence.

Metamemory and Reporting of Information

Most research on the predictors of eyewitness accuracy and on the eyewitness confidence-accuracy relation have focused mainly on identification settings, without reference to situations where eyewitness recall information about a crime. Witnesses' recollections of an event of interest can be critical in investigative processes and court decisions, including, for example, providing descriptions of a perpetrator or clarifying the facts of a crime incident. Situations in which an eyewitness is required to report information they remember are fundamentally different from situations in which an eyewitness is required

to recognize one specific target information (e.g., lineup identifications). In recognition tests, individuals are exposed to a stimulus and are required to judge whether they have previously perceived that stimulus or not. In free recall tests, individuals are required to freely report information they can remember about a target stimulus. Witnesses who recall information about an event can regulate their output in order to balance informativeness and accuracy, withholding details that fall below a pre-set criterion of probable accuracy (Koriat & Goldsmith, 1996; Sauer & Hope, 2016). This regulation process present in free recall often results in a stronger confidence-accuracy relation when compared to situations involving forced-responses or recognition paradigms (Allwood, Ask, & Granhag, 2005; Robinson, Johnson, & Herndon, 1997; Robinson, Johnson, & Robertson, 2000; Robinson & Johnson, 1996). However, at present it is unknown whether witnesses regulate their reporting of information (and associated confidence) based on internal cues related to memory self-efficacy.

In Experiment 6 (**Chapter 6**), we investigated this issue, expecting that witnesses would take into account their self-assessments of memory capacity to regulate the reporting of information in a free recall setting (Evans & Fisher, 2011). However, in this experiment we provide some of the first evidence that assessments of general memory self-efficacy are in fact unrelated to witnesses free recall completeness, accuracy and confidence. These results differ from those obtained in previous studies (Olsson & Juslin, 1999; Searcy, Bartlett, & Memon, 2000) and in other experiments in the current programme of work regarding the relation between memory self-efficacy and eyewitness identification performance (**Chapters 2 to 5**). Taken together, these observations seem to highlight important distinctions between recall and recognition tasks. Some theoretical frameworks propose that memory monitoring processes may be based on general beliefs of memory capacity (e.g., “I do not have a very good memory”) or on the learning experience (i.e., the experience of remembering the

learning material (Koriat, Nussinson, Bless, & Shaked, 2008). It may be the case that learning experience is more influential in free recall settings than general beliefs of memory capacity, so that witnesses prioritize their memory traces over feelings of memory capacity when disclosing information during free recall. In other words, two witnesses with very different opinions about their own memory self-efficacy may report the same amount and quality of information depending on their available memory traces. This finding further highlights that individuals can monitor their memories and control what they report in order to achieve an acceptable accuracy, regardless of internal cues and beliefs related to memory capacity.

Although we find no evidence that memory self-efficacy relates to eyewitness free recall performance, this relationship needs to be examined under more diverse or challenging conditions. We conducted one experiment (Experiment 6) in which participants completed a relatively easy free recall task, involving a short retention interval (5 minutes) between the event to be recalled and the recall test. It might be argued that in this situation participants were very likely to have strong memory traces of the to-be-remembered event at the time of the free recall test. Consequently, the presence of strong memory traces may have been the most important factor underlying recall accuracy, completeness and confidence, while the contribution of other factors such as memory self-efficacy may have been diminished. Future research should examine performance and associated relationships with recall tasks of varied difficulty by manipulating encoding and retrieval conditions, including the use of different stimuli or varied retention intervals. Further examination of the relationship between metamemory and free recall performance seems especially relevant given that previous studies have found an association between eyewitness accounts and other individual differences such as working memory capacity, intelligence and temperament (Chae & Ceci, 2005; Jaschinski & Wentura, 2002; Zhu et al., 2010). If more evidence is obtained confirming the absence of a relationship between MSE and eyewitness free recall, then it may be

unwarranted to discount an eyewitness account based on evaluations of memory self-efficacy. That is, an eyewitness may still provide an accurate and complete account even if it has been previously shown that the eyewitness distrusts their general memory ability.

Objective Tests of Memory Performance

Despite the potential contributions of memory self-efficacy as estimators of eyewitness performance, it might be argued that objective memory tests may be more informative with respect to eyewitness accuracy, confidence and confidence-accuracy relation. In fact, some studies have found that accuracy in different types of face memory tests are predictive of eyewitness identification accuracy (Bindemann et al., 2012; Morgan et al., 2007). In Experiments 1, 2 and 6, we examined the relationship between memory tests of different domains (i.e., free recall, cued recall, face recognition and general knowledge) and performance in identification tasks (Experiments 1 and 2, **Chapter 2**) and free recall tasks (Experiment 6, **Chapter 6**). By testing those relationships, we sought to elucidate whether or not individuals present similar memory accuracy and confidence across tests of different domains. In these experiments, it was expected that memory tests that were more similar in content and format to the eyewitness memory task (i.e., lineup identification or free recall) would be more predictive of eyewitness performance.

In Experiments 1 and 2 (**Chapter 2**), we found no evidence that performance on any of the objective memory tests used (i.e., general knowledge, cued recall, and face recognition) was related to performance on the eyewitness identification task and specifically, accuracy on that task. That is, it was not possible to infer an eyewitness's identification accuracy from his or her performance in other memory tests, even if the test was somewhat close in content and format to an eyewitness identification task (i.e., a face recognition test). In contrast, we found that eyewitness confidence on the identification test could be predicted

by the confidence expressed in different memory tests. This effect was also stronger for memory tests with more proximal domains, so that the strongest relationship observed was between the identification confidence and face recognition confidence, followed by cued recall confidence and general knowledge confidence. This pattern of results was observed in both Experiments 1 and 2 in **Chapter 2**. Similarly, in Experiment 6 (**Chapter 6**) we find that eyewitness free recall accuracy could only be predicted by accuracy in a cued recall test, while eyewitness free recall confidence could be predicted by confidence in free recall, cued recall and general knowledge tests.

In sum, these findings suggest that there is some stability in how individuals report confidence across different memory tests, with a closer correspondence for tests of similar domain. Some theoretical models of memory propose that memory consists of relatively independent systems (Baddeley, 2000; Tulving, 2007). These distinct systems can share basic features (e.g., the means of acquiring new information), but differ in other features (e.g., functions and underlying neural mechanisms (Schacter, Wagner, & Buckner, 2000; Squire, 2004). Some evidence suggests that an individual performance in a memory system may not necessarily transfer to performance in a different memory system (Conway et al., 2005; Redick et al., 2013). Overall our findings seem to reinforce this premise, given that in our experiments almost no stability was observed in accuracy across different memory tests. Interestingly, we do observe that confidence judgements are somewhat stable across different memory tasks. So, although individuals may have poorer or better performance depending on the memory task at hand, their confidence tends to remain similar across the different tasks (Bornstein & Zickafoose, 1999). This finding may be explained by theoretical frameworks proposing that confidence judgments are not only based on memory traces, but also on intrinsic cues related to memory capacity (Koriat, 2000). That is, individuals may present stable confidence judgements across different memory tasks because they hold stable

opinions and beliefs about their general memory performance. If this hypothesis is further corroborated, it may represent a foundation for assessments of confidence in objective memory tests to estimate witness's confidence in identification or free recall settings (Bornstein & Zickafoose, 1999).

Methodological Limitations

In order to best interpret the findings and implications of the current programme of work it is important to consider a number of methodological limitations in this set of experiments. First, it is important to note that across our experiments examining the relationship between metamemory and eyewitness memory performance there was a short retention interval between the event of interest (e.g., mock crime) and the memory test (e.g., lineup identification or free recall task). In most of the experiments (Experiments 1, 2, 4, 5 and 6) participants first watched the event of interest, then took part in a short distractor task (approximately 5 to 10 minutes) and then completed the memory task. Our main rationale for using a short retention interval was almost entirely pragmatic. We reasoned that a long retention interval between the target event and the memory task would greatly increase the practical difficulties in recruiting the sample sizes needed for each experiment. We also reasoned that the cleanest relational effects would be observed in the context of a short interval before much forgetting had taken place. However, eyewitness accounts in real world cases are often obtained after a longer retention interval reaching weeks or months in some cases (Behrman & Davey, 2001; Horry, Halford, Brewer, Milne, & Bull, 2014). Future studies examining the relation between self-assessments of metamemory and eyewitness memory performance may benefit from testing longer retention intervals. Indeed, it may be expected, for example, that internal cues related to memory self-efficacy may be more predictive of how witness's express confidence in identification tasks after a delay. The

rationale for this prediction is that memory traces associated to the perpetrators identity tend to fade as retention intervals increases (Palmer et al., 2013; Sauer et al., 2010), so witnesses may rely more on stable internal cues (e.g., memory self-efficacy for face recognition) to estimate their identification accuracy. It is also important to further examine the diagnostic value of metamemory assessments as estimators of eyewitness identification accuracy and longer periods of retention interval.

Another important limitation concerns the order of metamemory assessments and the eyewitness paradigms. In our experiments, we decided to conduct the metamemory assessments prior to the eyewitness paradigm, in either separate sessions (Experiments 1 and 2, **Chapter 2**) or in the same session (Experiments 4 and 5, **Chapters 4 and 5**). However, this order does not replicate what would occur in real case scenarios, given that metamemory assessments would most likely be obtained after an identification decision or an eyewitness account has already been obtained. In our experiments, we were mostly concerned with possible confounding variables affecting the metamemory assessments rather than any possible effects that conducting the metamemory assessments might have on eyewitness memory performance. With this reasoning we expected that completing an eyewitness paradigm would affect the assessments of memory self-efficacy to a greater extent than the assessments would affect eyewitness performance. Although this approach may be adequate for our current goals, it is still important to examine whether the findings obtained in our experiments can be replicated when obtaining metamemory assessments after the eyewitness paradigm has occurred.

It is also important to note that our materials had some degree of artificiality compared to what occurs in a real eyewitness context. For example, in our experiments we adopted mock-crime videos as the main to-be-remembered stimuli, instead of live simulations of a target event. Although some evidence suggests that eyewitness performance does not

differ between video and live interactions (Pozzulo, Crescini, & Panton, 2008), it is still important to examine if our findings can be replicated in more naturalistic contexts. We chose to use videos in our experiments due to pragmatic reasons, given that the use of live scenarios would be very resource intensive in order to achieve our desired sample sizes. With this approach, we were also able to use more diverse stimuli containing different targets and types of mock crime, which may allow us to better generalize our findings. Finally, by using mock-crime videos we were able to conduct our experiments on online settings, allowing us to reach a more diverse sample other than college students.

Implications of this Research

The criminal justice system relies heavily on eyewitness accounts and identifications, especially when other types of physical evidence are lacking. A notable issue involving eyewitness evidence is that retrieved memories may not always be accurate, and witnesses can misinterpret the facts of a crime or wrongly identify an innocent suspect, even if the witness provides an honest testimony. Psychological research has greatly advanced our understanding of factors that can contaminate witnesses memories, so that it is now possible to better evaluate the diagnostic value of this type of evidence in a case-by-base manner (Lindsay, Ross, Don Read, & Toglia, 2013; Wixted et al., 2018). However, confidently distinguishing accurate from inaccurate eyewitness evidence remains a major challenge because only a few effective postdictors of eyewitness performance have been identified so far. It has been extensively demonstrated, for example, that eyewitnesses confidence have a positive albeit not perfect relationship with identification accuracy (Palmer et al., 2013; Sauer et al., 2010; Wixted & Wells, 2017), and witnesses who are faster in their identifications tend to be more accurate than those who take a longer time to decide whether the perpetrator is present or not in a lineup (Sauerland & Sporer, 2007; Weber, Brewer, Wells, Semmler, &

Keast, 2004). The current thesis makes a unique contribution to this line of research by presenting evidence that some metamemory assessments can be useful in estimating identification accuracy and on elucidating how witnesses express their confidence.

Perhaps of most applied value, we present initial evidence that metamemory assessments specific to face recognition and person identification are predictive of eyewitness identification performance. Importantly, we find that metamemory assessments are predictive of identification performance even when biased lineups are presented to witnesses, a situation which often diminishes the diagnostic value of other postdictors of identification accuracy. Considering the limitations inherent in our experiments, it is too early to recommend the use of metamemory assessments to distinguish accurate from inaccurate identifications. Before applied use is warranted, it will be important to establish whether our findings can be replicated in more naturalistic settings, including metamemory assessments after the identification task has been completed. Nevertheless, our findings lay the foundation for further research examining this issue. Given the results obtained across our experiments, we highlight the importance of adopting valid metamemory assessments that are specifically related to the memory demands present in identification settings. This line of research would also greatly benefit from studies examining the predictive value of metamemory assessments obtained after the identification task has occurred. Additionally, it is important to consider the practical challenges of adopting such measures in practice, by possibly refining the assessments so that the same constructs can be measured with more concise scales. With further replication of these findings, metamemory assessments could be used in addition to other predictors of identification accuracy, such as confidence judgements, decision time and decision processes, in order to better inform decisions regarding eyewitness identification evidence.

One clear positive benefit of using valid metamemory assessments in practice is that it may assist the evaluation of eyewitness evidence accuracy. This in-depth evaluation of the diagnosticity of eyewitness evidence is important for investigators and triers of fact, resulting in better decisions regarding lines of investigation and the likely guilty of a suspect.

However, it must also be noted that metamemory assessments may also have negative effects in practice. That is because psychometric assessments may be constructed in numerous ways with varying degrees of validity. Therefore, it is crucial to consider adequate methodological approaches and standards of validity to build a solid foundation for the use of metamemory assessments in practice. In the current programme of work, we aimed to adopt the highest standards in scientific practice, including preregistrations of experiments in open science platforms and estimation of required sample sizes via power analysis. Although further replication of our findings is required, we provide a basis for future studies examining the utility of metamemory assessments in field work.

In addition to practical implications, the current thesis has important contributions to theory. We consistently observed some stability in confidence judgements across tests of different domains, supporting predictions that metamemory judgements are based partially on stable internal cues related to memory capacity (Nelson, 1990). In fact, we also obtained extensive evidence of a direct relation between valid measures of memory self-efficacy and confidence judgements in eyewitness identification and free recall tasks. These findings support the importance of dispositional factors in how one approaches memory confidence judgements. That is, individuals do seem to use a variety of external and internal cues when assessing how likely a specific information is recalled or recognized (Koriat, 2000; Leippe & Eisenstadt, 2014). We further contribute to theoretical developments on metamnemonic judgements by demonstrating that domain specific memory self-efficacy is more strongly related to confidence judgements, compared to general assessments of memory self-efficacy.

Therefore, it is important that theoretical models of confidence judgements take into account that individuals have distinct self-efficacy for different memory domains, and confidence judgements will be mostly related to the memory self-efficacy representing the task at hand.

Future Research Directions

There are several interesting routes to further extend the work described in this thesis. In order to address the limitations in our experiments, future studies could: i) adopt varied degrees of retention interval in their paradigms; ii) test metamemory assessments after exposure to the target event; and iii) adopt naturalistic target events such as staged mock-crimes or real interactions (e.g., Valentine & Mesout, 2009). One important line of research to be explored is how metamemory relates to different types or categories of witness, in particular the performance of older witnesses. Senior citizens are increasingly active in society, which in turn means they are more likely to become involved in the Criminal Justice System by being victims or witnesses of accidents and crimes (Rothman, Dunlop, & Pamela, 2004). It is important to examine whether our findings replicate or differ when assessing elder witnesses, given that both memory and meta-mnemonic mechanisms are age related (Grady & Craik, 2000). There is a general agreement that not only memory, but also metamemory processes declines from early to late adulthood (Henkel, Toglia, Ross, Pozzulu, & Pica, 2014). Therefore, it remains to be examined whether the results obtained in the current programme of research would replicate in the elderly population.

Another line of research yet to be explored is the relationship between metamemory and susceptibility to memory misinformation. Several studies point to a relation between eyewitness susceptibility to misinformation and individual differences such as working memory capacity, intelligence and temperament (Chae & Ceci, 2005; Jaschinski & Wentura,

2002; Zhu et al., 2010). Therefore, it remains to be examined whether self-ratings of memory capacity are related to resistance or susceptibility to misinformation.

Finally, it would appear that there are pervasive gaps in our knowledge regarding the structure and dimensionality of metamemory constructs. Metamemory instruments and paradigms have shown predictive value for a variety of cognitive processes, but the conceptualization and dimensionality of metamemory is a critical point of debate that requires further investigation (Dunlosky & Bjork 2008). It is still unclear, for example, how many distinguishable dimensions are associated to metamemory (e.g., contentment, ability, strategies), and how those dimensions relate to each other. We recommend the examination of the multidimensional structure of metamemory by conducting exploratory and confirmatory factor analysis on diverse metamemory self-assessment tools (i.e., Troyer and Rich 2002; Squire et al. 1979; Bergen et al. 2010). A better understanding on the dimensions of metamemory may results in more complete and accurate theories that can generate precise testable hypotheses.

Conclusion

In the current programme of work, our efforts were focused towards elucidating the relation between metamemory and eyewitness identification accuracy, confidence and confidence-accuracy relation. With this set of experiments, we advance the thesis that in order to improve or better estimate eyewitness memory accuracy, it is crucial to understand how people monitor and evaluate their own memory cognitive processes. Specifically, we contribute to our understanding of the relation between self-assessments of memory-efficacy and eyewitnesses' performance. With this new knowledge, we provide a solid foundation for research investigating the relation between metamemory assessments and eyewitness memory accuracy and confidence in the future. Although still in its infancy, this line of research has important implications for the understanding and application of postdictors of

Chapter 7

eyewitness accuracy, ultimately contributing to a better evaluation of eyewitness evidence in the criminal justice system.

REFERENCES

- Allwood, C. M., & Jonsson, A. C. (2005). The effects of source and type of feedback on child witnesses' metamemory accuracy. *Applied Cognitive Psychology, 19*, 331-344. doi: 10.1002/acp.1071
- Allwood, C. M., Ask, K., & Granhag, P. A. (2005). The Cognitive Interview: Effects on the realism in witnesses' confidence in their free recall. *Psychology, Crime & Law, 11*, 183-198. doi: 10.1080/10683160512331329943
- Andersen, S. M., Carlson, C. A., Carlson, M. A., & Gronlund, S. D. (2014). Individual differences predict eyewitness identification performance. *Personality and Individual Differences, 60*, 36-40. doi: 10.1016/j.paid.2013.12.011
- Aslan, A., & Bäuml, K. H. T. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*, 264-269. doi:10.1037/a0021324
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423. doi: 10.1016/S1364-6613(00)01538-2
- Baldassari, M. J., Kantner, J., & Lindsay, D. S. (2019). The importance of decision bias for predicting eyewitness lineup choices: Toward a Lineup Skills Test. *Cognitive Research: Principles and Implications, 4*, 2. doi: 10.1186/s41235-018-0150-3
- Bang, D., Aitchison, L., Moran, R., Herce Castanon, S., Rafiee, B., Mahmoodi, A., ... Summerfield, C. (2017). Confidence matching in group decision-making. *Nature Human Behaviour, 1*(0117), 1-17. doi: 10.1038/s41562-017-0117
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. Retrieved from <http://arxiv.org/abs/1406.5823>
- Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological bulletin, 137*, 211-241. doi:10.1037/a0022106
- Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An

References

- archival analysis. *Law and Human Behavior*, 25, 475–491. doi: 10.1023/A:1012840831846
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology*, 20, 115–129. doi: 10.1002/acp.1171
- Bergen, S., Brands, I., & Jelicic, M. (2010). Assessing trait memory distrust: Psychometric properties of the Squire Subjective Memory Questionnaire. *Legal and Criminological Psychology*, 15, 373–384. doi: 10.1348/135532509X471960
- Bergold, A. N., & Heaton, P. (2018). Does filler database size influence identification accuracy? *Law and Human Behavior*, 42, 227–243. doi: 10.1037/lhb0000289
- Berry, J. M. (1999). Memory self-efficacy in its social cognitive context. In T. M. Hess & F. Blanchard-Fields (Eds.), *Social Cognition and Aging* (pp. 69–96). London, UK: Academic Press.
- Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, 41, 897–903. doi: 10.3758/s13421-013-0307-8
- Bindemann, M., Attard, J., & Johnston, R. A. (2014). Perceived ability and actual recognition accuracy for unfamiliar and famous faces. *Cogent Psychology*, 1(986903), 1–15. doi:10.1080/23311908.2014.986903
- Bindemann, M., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of*

- Experimental Psychology. Applied*, 18, 277–291. doi: 10.1037/a0029635
- Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification predict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, 1, 96–103. doi:10.1016/j.jarmac.2012.02.001
- Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012/6). Individual differences in face identification predict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, 1, 96–103. doi: 10.1016/j.jarmac.2012.02.001
- Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLoS One*, 11, e0148148. doi:10.1371/journal.pone.0148148
- Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities. *Quarterly Journal of Experimental Psychology*, 72, 872–881. doi: 10.1177/1747021818776145
- Bobak, A. K., Pampoulov, P., & Bate, S. (2016). Detecting superior face recognition skills in a large sample of young British adults. *Frontiers in Psychology*, 7(1378), 1–11. doi: 10.3389/fpsyg.2016.01378
- Bornstein, B. H., & Zickafoose, D. J. (1999). “I know I know it, I know I saw it”: The stability of the confidence--accuracy relationship across domains. *Journal of Experimental Psychology. Applied*, 5, 76–88. doi: 10.1037/1076-898X.5.1.76
- Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, 72, 691–695. doi: 10.1037/0021-9010.72.4.691
- Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *The*

References

- Journal of Applied Psychology*, 87, 112–120. doi: 10.1037//0021-9010.87.1.112
- Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior*, 26, 353–364. doi: 10.1023/A:1015380522722
- Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, 12(1), 11–30. doi: 10.1037/1076-898X.12.1.11
- Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: the effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied*, 8(1), 44–56. doi:10.1037/1076-898X.8.1.44
- Brewer, N., Weber, N., & Semmler, C. (2013). A role for theory in eyewitness identification research. In R. C., Lindsay, D. F., Ross., J. D., Read, & M. P. Toglia (Eds.), *The Handbook of Eyewitness Psychology: Volume II: Memory for People* (pp. 201–218). New York, NY: Routledge.
- Brewer, W. F., & Sampaio, C. (2012). The metamemory approach to confidence: A test using semantic memory. *Journal of Memory and Language*, 67, 59–77. doi: 10.1016/j.jml.2012.04.002
- Brewer, W. F., Sampaio, C., & Barlow, M. R. (2005). Confidence and accuracy in the recall of deceptive and nondeceptive sentences. *Journal of Memory and Language*, 52, 618–627. doi: 10.1016/j.jml.2005.01.017
- Bruck, M., & Melnyk, L. (2004). Individual differences in children's suggestibility: A review and synthesis. *Applied Cognitive Psychology*, 18, 947–996. doi: 10.1002/acp.1070
- Buratti, S., Allwood, C. M., & Johansson, M. (2014). Stability in the metamemory realism of

- eyewitness confidence judgments. *Cognitive Processing*, 15, 39–53. doi: 10.1007/s10339-013-0576-y
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer. doi: 10.1007/b97636
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, 42, 286–291. doi: 10.3758/BRM.42.1.286
- Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic bulletin & review*, 7, 26–48.
- Cabeza, R., & Moscovitch, M. (2013). Memory systems, processing modes, and components: functional neuroimaging evidence. *Perspectives on Psychological Science*, 8, 49–55. doi: 10.1177/1745691612469033
- Cavanaugh, J. C., & Green, E. E. (1990). I believe, therefore I can: Self-efficacy beliefs in memory aging. In E. A. Lovelace (Ed.), *Advances in Psychology* (Vol. 72, pp. 189–230). [https://doi.org/10.1016/S0166-4115\(08\)60788-8](https://doi.org/10.1016/S0166-4115(08)60788-8)
- Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: A critical examination. *Child Development*, 53, 11–28. doi:10.2307/1129635
- Chae, Y., & Ceci, S. J. (2005). Individual differences in children's recall and suggestibility: the effect of intelligence, temperament, and self-perceptions. *Applied Cognitive Psychology*, 19, 383–407. doi: 10.1002/acp.1094
- Chae, Y., & Ceci, S. J. (2005). Individual differences in children's recall and suggestibility: the effect of intelligence, temperament, and self-perceptions. *Applied Cognitive Psychology*, 19, 383–407. doi: 10.1002/acp.1094
- Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & Cognition*, 22, 273–280. doi: 10.3758/BF03200854

References

- Charman, S. D., Wells, G. L., & Joy, S. W. (2011). The dud effect: Adding highly dissimilar fillers increases confidence in lineup identifications. *Law and Human Behavior*, 35, 479–500. doi:10.1007/s10979-010-9261-1
- Chua, E. F., Hannula, D. E., & Ranganath, C. (2012). Distinguishing highly confident accurate and inaccurate memory: Insights about relevant and irrelevant influences on memory confidence. *Memory*, 20, 48–62. doi: 10.1080/09658211.2011.633919
- Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2006). Understanding metamemory: Neural correlates of the cognitive process and subjective level of confidence in recognition memory. *Neuroimage*, 29, 1150–1160. doi: 10.1016/j.neuroimage.2005.09.058
- Clark, S. E. (1997). A familiarity-based account of confidence–accuracy inversions in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(1), 232–238. doi: 10.1037/0278-7393.23.1.232
- Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science*, 27, 1227–1239. doi: 10.1177/0956797616655789
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786. doi: 10.3758/BF03196772
- Crook, T. H., & Larrabee, G. J. (1990). A self-rating scale for evaluating memory in everyday life. *Psychology and Aging*, 5, 48–57. doi: 10.1037//0882-7974.5.1.48
- Dahl, M., Allwood, C. M., Scimone, B., & Rennemark, M. (2015). Old and very old adults as witnesses: event memory and metamemory. *Psychology, Crime & Law*, 21, 764–775. doi: 10.1080/1068316X.2015.1038266
- Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything

- about their relationship?. *Law and Human Behavior*, 4, 243-260. doi: 10.1007/BF01040617
- Dixon, R. A., Hultsch, D. F., & Hertzog, C. (1988). The Metamemory in Adulthood (MIA) questionnaire. *Psychopharmacology Bulletin*, 24, 671-688.
- Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory*, 26, 741-750. doi: 10.1080/09658211.2017.1404111
- Douglass, A. B., & Jones, E. E. (2013). Confidence inflation in eyewitnesses: Seeing is not believing. *Legal and Criminological Psychology*, 18, 152-167. doi: 10.1111/j.2044-8333.2011.02031.x
- Douglass, A. B., & Steblay, N. (2006). Memory distortion in eyewitnesses: a meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology*, 20, 859-869. doi: 10.1002/acp.1237
- Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology*, 106, 433-445. doi: 10.1111/bjop.12103
- Dunlosky, J., & Bjork, R. A. (2008). *Handbook of Metamemory and Memory*. Psychology Press.
- Dunlosky, J., & Bjork, R. A. (2008). The integrated nature of metamemory and memory. In J., Dunlosky, & R. A., Bjork (Eds.), *Handbook of Metamemory and Memory*, (pp. 11-28). New York, NY: Psychology Press.
- Dunlosky, J., & Bjork, R. A. (2013). The integrated nature of metamemory and memory. In J. Dunlosky, & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 11-28). London: Psychology Press.
- Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A

References

- componential analysis of learning about strategy effectiveness from task experience. *Psychology and aging*, 15, 462-474. doi: 10.1037//0882-7974.15.3.462
- Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory. In J. Dunlosky & S. K. Tauber (Ed.), *The Oxford Handbook of Metamemory*. New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780199336746.013.14
- Dunn, T. J., Baguley, T., & Brunsten, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 399–412. doi:10.1111/bjop.12046
- Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, 67, 818-835. doi: 10.1037/0022-3514.67.5.818
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12, 83–87. doi: 10.1111/1467-8721.01235
- Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology*, 25, 501–508. doi:10.1002/acp.1722
- Finney, S. J., & DiStefano, C. (2013). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modelling: A second course* (pp. 439-492). Charlotte, NC: Information Age Publishing.
- Fitzgerald, R. J., Oriet, C., & Price, H. L. (2015). Suspect filler similarity in eyewitness lineups: A literature review and a novel methodology. *Law and Human Behavior*, 39, 62–74. doi: 10.1037/lhb0000095
- Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler

- similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, 19, 151-164. doi: 10.1037/a0030618
- Fort, I., Adoul, L., Holl, D., Kaddour, J., & Gana, K. (2004). Psychometric properties of the French version of the Multifactorial Memory Questionnaire for adults and the elderly. *Canadian Journal on Aging*, 23, 347-357. doi: <https://doi.org/10.1353/cja.2005.0020>
- Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 43, 680–693. doi:10.1037/xlm0000332
- Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, 9, 641-651. doi: 10.1177/1745691614551642
- Ghetti, S., Lyons, K. E., Lazzarin, F., & Cornoldi, C. (2008). The development of metamemory monitoring during retrieval: The case of memory strength and memory absence. *Journal of Experimental Child Psychology*, 99, 157–181. doi:10.1016/j.jecp.2007.11.001
- Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General*, 131, 73-95. doi: 10.1037/0096-3445.131.1.73
- Goldsmith, M., Pansky, A., & Koriat, A. (2014). Metacognitive control of memory reporting. In T.J. Perfect, & D.S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 481-500). Sage: London.
- Grady, C. L., & Craik, F. I. (2000). Changes in memory processing with age. *Current Opinion in Neurobiology*, 10, 224-231. doi: 10.1016/S0959-4388(00)00073-8
- Granhag, P. A. (1997). Realism in eyewitness confidence as a function of type of event witnessed and repeated recall. *The Journal of Applied Psychology*, 82, 599–613. doi:

References

- 10.1037//0021-9010.82.4.599
- Gray, K. L. H., Bird, G., & Cook, R. (2017). Robust associations between the 20-item prosopagnosia index and the Cambridge Face Memory Test in the general population. *Royal Society Open Science*, 4, 160923. doi: 10.1098/rsos.160923
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7, 493–498. doi: 10.1111/2041-210X.12504
- Guerrero Sastoque, L., Bouazzaoui, B., Burger, L., Froger, C., Isingrini, M., & Taconnat, L. (2019). Optimizing memory strategy use in young and older adults: The role of metamemory and internal strategy use. *Acta Psychologica*, 192, 73–86. doi:10.1016/j.actpsy.2018.11.002
- Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). *Handbook of metacognition in education*. Routledge.
- Henkel, L. A., Tolia, M. P., Ross, D. F., Pozzulu, J., & Pica, E. (2014). Memory trust and distrust in elderly eyewitnesses: To what extent do older adults doubt their memories. In M. P., Tolia, D. F., Ross, J., Pozzulo, & E. Pica (Eds.), *The elderly eyewitness in court* (pp. 232-262). New York, NY: Psychology Press.
- Hertzog, C. (2002). Metacognition in older adults: Implications for application. In T. J., Perfect, & B. L. Schwartz (Eds.), *Applied Metacognition*, (pp. 169–196). Cambridge, UK: Cambridge University Press.
- Hertzog, C., & Dixon, R. A. (1994). Metacognitive development in adulthood and old age. In J. Metcalfe, & A. P. Shimamura (Eds.), *Metacognition: Knowing about Knowing* (pp. 227–251). Cambridge: MIT press.
- Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, 5,

- 215–227. doi: 10.1037/0882-7974.5.2.215
- Higham, P. A., & Vokey, J. R. (2000). Judgment heuristics and recognition memory: Prime identification and target-processing fluency. *Memory & Cognition*, 28, 574–584. doi: 10.3758/BF03201248
- Horry, R., Halford, P., Brewer, N., Milne, R., & Bull, R. (2014). Archival analyses of eyewitness identification test outcomes: what can they tell us about eyewitness memory? *Law and Human Behavior*, 38, 94–108. doi: 10.1037/lhb0000060
- Howie, P., & Roebbers, C. M. (2007). Developmental progression in the confidence-accuracy relationship in event recall: insights provided by a calibration perspective. *Applied Cognitive Psychology*, 21, 871–893. doi: 10.1002/acp.1302
- Dunlosky, J., & Bjork, R. A., (2013), *Handbook of metamemory and memory*. New York, NY: Psychology Press.
- Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology. General*, 118, 126–135. doi: 10.1037/0096-3445.118.2.126
- Jaschinski, U., & Wentura, D. (2002). Misleading postevent information and working memory capacity: an individual differences approach to eyewitness memory. *Applied Cognitive Psychology*, 16, 223–231. doi: 10.1002/acp.783
- Jonsson, A.-C., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences*, 34, 559–574. doi: 10.1016/S0191-8869(02)00028-4
- Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3, 759–763. doi: 10.1038/77664
- Kebbell, M. R., Wagstaff, G. F., & Covey, J. A. (1996). The influence of item difficulty on the relationship between eyewitness confidence and accuracy. *British Journal of Psychology*,

References

- 87, 653-662. doi: 10.1111/j.2044-8295.1996.tb02614.x
- Kelemen, W. L. (2000). Metamemory cues and monitoring accuracy: Judging what you know and what you will know. *Journal of Educational Psychology*, 92, 800-810. doi:10.1037/0022-0663.92.4.800
- Key, K. N., Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Cash, D. K., & Lane, S. (2017). Line-up fairness affects postdictor validity and “don’t know” responses. *Applied Cognitive Psychology*, 31, 59–68. doi: 10.1002/acp.3302
- Kliegel, M., & Jäger, T. (2006). Can the prospective and retrospective memory questionnaire (PRMQ) predict actual prospective memory performance? *Current Psychology*, 25, 182–191. doi: 10.1007/s12144-006-1002-8
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review* 100, 609-639. doi:10.1037/0033-295X.100.4.609
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349-370. doi: 10.1037/0096-3445.126.4.349
- Koriat, A. (2000). The feeling of knowing: some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, 9, 149–171. doi: 10.1006/ccog.2000.0433
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103, 490–517. doi: 10.1037/0033-295X.103.3.490
- Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: the role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133, 643-656. doi: 10.1037/0096-3445.133.4.643
- Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy.

- Annual Review of Psychology*, 51, 481–537. doi: 10.1146/annurev.psych.51.1.481
- Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, 135, 36–69. doi: 10.1037/0096-3445.135.1.36
- Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky, & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 117–136). London: Psychology Press.
- Kruschke, J. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press. London: UK.
- Lane, C. J., & Zelinski, E. M. (2003). Longitudinal hierarchical linear models of the memory functioning questionnaire. *Psychology and Aging*, 18, 38–53. doi: 10.1037/0882-7974.18.1.38
- Leippe, M. R., & Eisenstadt, D. (2014). Eyewitness confidence and the confidence-accuracy relationship in memory for people. R. C. Lindsay, D. F. Ross, J. D. Read, & M. M. P. Toglia (Eds.), *The handbook of eyewitness psychology: volume ii: memory for people* (pp. 377–425). London, UK: Psychology Press.
- Leippe, M. R., Eisenstadt, D., & Rauch, S. M. (2009). Cueing confidence in eyewitness identifications: Influence of biased lineup instructions and pre-identification memory feedback under varying lineup conditions. *Law and Human Behavior*, 33, 194–212. doi: 10.1007/s10979-008-9135-y
- Leippe, M. R., Eisenstadt, D., Rauch, S. M., & Stambush, M. A. (2006). Effects of social-comparative memory feedback on eyewitnesses' identification confidence, suggestibility, and retrospective memory reports. *Basic and Applied Social Psychology*, 28, 201–220.

References

- doi: 10.1207/s15324834basp2803_1
- Liebman, J. I., McKinley-Pace, M. J., Leonard, A. M., Sheesley, L. A., Gallant, C. L., Renkey, M. E., & Lehman, E. B. (2002). Cognitive and psychosocial correlates of adults' eyewitness accuracy and suggestibility. *Personality and Individual Differences*, 33, 49–66. doi:10.1016/S0191-8869(01)00135-0
- Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science*, 9, 215–218. doi: 10.1111/1467-9280.00041
- Lindsay, R. C. L., Ross, D. F., Don Read, J., & Toglia, M. P. (2013). *The Handbook of Eyewitness Psychology: Volume II: Memory for People*. London, UK: Psychology Press.
- Livingston, L. A., & Shah, P. (2018). People with and without prosopagnosia have insight into their face recognition ability. *Quarterly Journal of Experimental Psychology*, 71, 1260–1262. doi: 10.1080/17470218.2017.1310911
- Manning, W. G., Basu, A., & Mullahy, J. (2005). Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*, 24, 465–488. doi: 10.1016/j.jhealeco.2004.09.011
- Mansour, J. K., Beaudry, J. L., & Lindsay, R. C. L. (2017). Are multiple-trial experiments appropriate for eyewitness identification studies? Accuracy, choosing, and confidence across trials. *Behavior Research Methods*, 49, 2235–2254. doi: 10.3758/s13428-017-0855-0
- Meinhardt, B., Persike, M., & Meinhardt, G. (2014). Holistic processing and reliance on global viewing strategies in older adults' face perception. *Acta Psychologica*, 151, 155-163. doi:10.1016/j.actpsy.2014.06.001
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49, 270–291. doi: 10.1037/a0028228

- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer”: Evidence from a meta-analytic review. *Perspectives on Psychological Science, 11*, 512–534. doi: 10.1177/1745691616635612
- Memon, A., Havard, C., Clifford, B., Gabbert, F., & Watt, M. (2011). A field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. *Psychology, Crime & Law, 17*, 711–729. doi: 10.1080/10683160903524333
- Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94*, 339–354. doi: 10.1348/000712603767876262
- Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: stability and generality in the learning process and predictive validity for learning outcome. *Memory & Cognition, 38*, 441–451. doi: 10.3758/MC.38.4.441
- Metcalfe, J., & Shimamura A. P. (1994), *Metacognition: Knowing about knowing* (pp. 227–251). Cambridge, Massachusetts: MIT press.
- Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*, 159–163. doi: 10.1111/j.1467-8721.2009.01628.x
- Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 19*, 851–861. doi: 10.1037/0278-7393.19.4.851
- Mickes, L. (2015). Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4*, 93–102. doi: 10.1016/j.jarmac.2015.01.003
- Mitchell, A. J. (2008). The clinical significance of subjective memory complaints in the

References

- diagnosis of mild cognitive impairment and dementia: A meta-analysis. *International Journal of Geriatric Psychiatry*, 23, 1191-1202. doi:10.1002/gps.2053
- Morgan, C. A., Hazlett, G., Baranoski, M., Doran, A., Southwick, S., & Loftus, E. (2007). Accuracy of eyewitness identification is significantly associated with performance on a standardized test of face recognition. *International Journal of Law and Psychiatry*, 30, 213–223. doi: 10.1016/j.ijlp.2007.03.005
- Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*, 20, 378–384. doi: 10.3758/s13423-012-0343-6
- Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 125–173). doi: 10.1016/S0079-7421(08)60053-5
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. *Metacognition: Knowing about Knowing*, 1–25. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1-26). Cambridge, Massachusetts: MIT press.
- Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy–confidence relation in eyewitness memory. *Applied Cognitive Psychology*, 20, 973–985. doi: 10.1002/acp.1263
- Odinot, G., Wolters, G., & van Giezen, A. (2013). Accuracy, confidence and consistency in repeated recall of events. *Psychology, Crime & Law*, 19, 629–642. doi: 10.1080/1068316X.2012.660152
- Olsson, N., & Juslin, P. (1999). Can self-reported encoding strategy and recognition skill be diagnostic of performance in eyewitness identifications? *The Journal of Applied Psychology*, 84, 42–49. doi: 10.1037/0021-9010.84.1.42
- Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor

- analysis: What it is and why it makes your analysis better. *Practical Assessment, Research & Evaluation*, 17, 1-8. ISSN: 1531-7714
- Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied*, 19, 55-71. doi: 10.1037/a0031602
- Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology review*, 15, 105-130. doi:10.1007/s11065-005-7091-6
- Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law*, 1, 817-845. doi: 10.1037/1076-8971.1.4.817
- Perfect, T. J. (2004). The role of self-rated ability in the accuracy of confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology*, 18, 157-168. doi: 10.1002/acp.952
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539–569. doi:10.1146/annurev-psych-120710-100452
- Pozzulo, J. D., Crescini, C., & Panton, T. (2008). Does methodology matter in eyewitness identification research? The effect of live versus video exposure on eyewitness identification accuracy. *International journal of law and psychiatry*, 31, 430-437. doi: 10.1016/j.ijlp.2008.08.006
- Prims, J., & Motyl, M. (2018). A tool for detecting low quality data in internet research. GitHub: <https://github.com/SICLab/detecting-bots>
- R Core Team (2019). R: *A language and environment for statistical computing*. R Foundation

References

- for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology. General*, 142, 359–379. doi: 10.1037/a0029082
- Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience*, 139(1), 5–21. doi:10.1016/j.neuroscience.2005.12.061
- Rickenbach, E. H., Agrigoroaei, S., & Lachman, M. E. (2015). Awareness of memory ability and change: (in)accuracy of memory self-assessments in relation to performance. *Journal of Population Ageing*, 8, 71–99. doi: 10.1007/s12062-014-9108-5
- Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology. Applied*, 6, 207–221. doi: 10.1037/1076-898X.6.3.207
- Rothman, M. B., Dunlop, B. D., & Pamela Entzel, J. D. (2004). *Elders, crime, and the criminal justice system: Myth, perceptions, and reality in the 21st century*. Springer Publishing Company.
- Russ, A. J., Sauerland, M., Lee, C. E., & Bindemann, M. (2018). Individual differences in eyewitness accuracy across multiple lineups of faces. *Cognitive Research: Principles and Implications*, 3, 30. doi: 10.1186/s41235-018-0121-8
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic bulletin & review*, 16, 252-257. doi:10.3758/PBR.16.2.252
- Saraiva, R. B., Van Boeijen, I. M., Hope, L., Horselenberg, R., & Van Koppen, P. (2019b, July 5). Using general and eyewitness-specific metamemory assessments to estimate

- performance in multiple identifications. <https://doi.org/10.31219/osf.io/pnhm3>
- Saraiva, R. B., Van Boeijen, I. M., Hope, L., Horselenberg, R., Sauerland, M., & Van Koppen, P. (2019, May 20). Development and validation of the Eyewitness Metamemory Scale. doi: 10.31234/osf.io/mjfea
- Saraiva, R. B., Van Boeijen, I. M., Hope, L., Sauerland, M., Horselenberg, R., & Van Koppen, P. (2019a, July 5). Eyewitness metamemory predicts identification performance in biased and unbiased lineups. <https://doi.org/10.31219/osf.io/tay75>
- Sauer, J. D., Brewer, N., & Wells, G. L. (2008). Is there a magical time boundary for diagnosing eyewitness identification accuracy in sequential line-ups? *Legal and Criminological Psychology*, 13, 123–135. doi:10.1348/135532506X159203
- Sauer, J., & Hope, L. (2016). The effects of divided attention at study and reporting procedure on regulation and monitoring for episodic recall. *Acta Psychologica*, 169, 143–156. doi: 10.1016/j.actpsy.2016.05.015
- Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence–accuracy relationship for eyewitness identification. *Law and Human Behavior*, 34, 337–347. doi: 10.1007/s10979-009-9192-x
- Sauer, J., Palmer, M. A., & Brewer, N. (in press). Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychology Public Policy and Law*. doi: 10.1037/law0000203
- Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law*, 6, 611–625. doi: 10.1080/10683160701264561
- Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology. Applied*, 15, 46–62. doi: 10.1037/a0014560

References

- Sauerland, M., Sagana, A., Sporer, S. L., & Wixted, J. T. (2018). Decision time and confidence predict choosers' identification performance in photographic showups. *PloS One*, 13(1), e0190416. doi: 10.1371/journal.pone.0190416
- Schacter, D. L., Wagner, A. D., & Buckner, R. L. (2000). Memory systems of 1999. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 627-643). New York: Oxford University Press.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modelling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99, 323–338. doi:10.3200/JOER.99.6.323-338
- Searcy, J. H., Bartlett, J. C., Memon, A., & Swanson, K. (2001). Aging and lineup performance at long retention intervals: Effects of metamemory and context reinstatement. *Journal of Applied Psychology*, 86, 207-214. doi: 10.1037/0021-9010.86.2.207
- Searcy, J., Bartlett, J. C., & Memon, A. (2000). Influence of post-event narratives, line-up conditions and individual differences on false identification by young and older eyewitnesses. *Legal and Criminological*, 5, 219-235. doi: 10.1348/135532500168100
- Seeman, T. E., Rodin, J., & Albert, M. (1993). Self-efficacy and cognitive performance in high-functioning older individuals: MacArthur studies of successful aging. *Journal of Aging and Health*, 5, 455–474. doi: 10.1177/089826439300500403
- Seeman, T., McAvay, G., Merrill, S., Albert, M., & Rodin, J. (1996). Self-efficacy beliefs and change in cognitive performance: MacArthur studies on Successful Aging. *Psychology and Aging*, 11, 538-551. doi: 10.1037/0882-7974.11.3.538
- Semmler, C., & Brewer, N. (2006). Postidentification feedback effects on face recognition confidence: Evidence for metacognitive influences. *Applied Cognitive Psychology*, 20, 895–916. doi: 10.1002/acp.1238

- Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. *Journal of Applied Psychology, 89*, 334-346. doi: 10.1037/0021-9010.89.2.334
- Shah, P., Sowden, S., Gaule, A., Catmur, C., & Bird, G. (2015). The 20-item prosopagnosia index (PI20): Relationship with the Glasgow face-matching test. *Royal Society Open Science, 2*(11), 150305. doi: 10.1098/rsos.150305
- Shaw III, J. S., & Zerr, T. K. (2003). Extra effort during memory retrieval may be associated with increases in eyewitness confidence. *Law and Human Behavior, 27*, 315-329. doi: 10.1023/A:1023487908640
- Shaw, J. S., & Zerr, T. K. (2003). Extra effort during memory retrieval may be associated with increases in eyewitness confidence. *Law and Human Behavior, 27*, 315-329. doi: 10.1023/A:1023487908640
- Shimamura, A. P. (2008). A neurocognitive approach to metacognitive monitoring and control. In J. Dunlosky, & R.A. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 373-390). Psychology Press: New York.
- Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do “brain-training” programs work? *Psychological Science in the Public Interest, 17*, 103-186. doi: 10.1177/1529100616661983
- Smith, A. M., Wilford, M. M., Quigley-McBride, A., & Wells, G. L. (2019). Mistaken eyewitness identification rates increase when either witnessing or testing conditions get worse. *Law and Human Behavior, 43*, 358-368. doi: 10.1037/lhb0000334
- Smith, S. M., Lindsay, R., & Pryke, S. (2001). Postdictors of eyewitness errors: Can false identifications be diagnosed in the cross-race situation? *Psychology, Public Policy, and Law, 7*, 153-159. doi: 10.1037/1076-8971.7.1.153
- Snijders, T. A. B. (2005). Power and sample size in multilevel modeling. *Encyclopedia of*

References

- Statistics in Behavioral Science*, 3, 1570–1573. doi: 10.1002/0470013192.bsa492
- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *The Journal of Applied Psychology*, 78, 22–33. doi: 10.1037/0021-9010.78.1.22
- Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, 118, 315–327. doi: 10.1037/0033-2909.118.3.315
- Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82, 171–177. doi: 10.1016/j.nlm.2004.06.005
- Squire, L. R., Wetzel, C. D., & Slater, P. C. (1979). Memory complaint after electroconvulsive therapy: assessment with a new self-rating instrument. *Biological Psychiatry*, 14, 791–801.
- Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence*, 28, 121–143. doi: 10.1016/S0160-2896(99)00033-1
- Stebly, N. K., & Dysart, J. E. (2016). Repeated eyewitness identification procedures with the same suspect. *Journal of Applied Research in Memory and Cognition*, 5, 284–289. doi: 10.1016/j.jarmac.2016.06.010
- Technical Working Group for Eyewitness Evidence. (2003). Eyewitness evidence: A trainer's manual for law enforcement. Washington, DC: National Institute of Justice.
- Thomaz, C. E., & Giraldi, G. A. (2010). A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing*, 28, 902–913. doi: 10.1016/j.imavis.2009.11.005
- Tonković, M., & Vranić, A. (2011). Self-evaluation of memory systems: Development of the questionnaire. *Aging & Mental Health*, 15, 830–837. doi:10.1080/13607863.2011.569483
- Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human*

- Behavior*, 22(2), 217–237. <https://doi.org/10.1023/A:1025746220886>
- Tredoux, C. G. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology*, 13, 9–26. doi: 10.1002/(SICI)1099-0720(199911)13:1+<S9::AID-ACP634>3.0.CO;2-1
- Tredoux, C. G. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology. Applied*, 8, 180–193. doi: 10.1037/1076-898X.8.3.180
- Troyer, A. K., & Rich, J. B. (2002). Psychometric properties of a new metamemory questionnaire for older adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 57, 19–27. doi:10.1093/geronb/57.1.P19
- Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological methods*, 6, 371–386. doi: 10.1037//1082-989X.6.4.371
- Tryon, W. W., & Lewis, C. (2008). An inferential confidence interval method of establishing statistical equivalence that corrects Tryon's (2001) reduction factor. *Psychological methods*, 13, 272–277. doi: 10.1037/a0013158
- Tulving, E. (2007). Are there 256 different kinds of memory. In Nairne, J. S. (Ed.). *The foundations of remembering: Essays in honor of Henry L. Roediger, III*. London: Psychology Press.
- Valentijn, S. A., Hill, R. D., Van Hooren, S. A., Bosma, H., Van Boxtel, M. P., Jolles, J., & Ponds, R. W. (2006). Memory self-efficacy predicts memory performance: Results from a 6-year follow-up study. *Psychology and Aging*, 21, 165–172. doi: 10.1037/0882-7974.21.2.165
- Valentine, T., & Mesout, J. (2009). Eyewitness identification under stress in the London

References

- Dungeon. *Applied Cognitive Psychology*, 23, 151-161. doi: 10.1002/acp.1463
- Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology*, 17, 969–993. doi: 10.1002/acp.939
- van Bergen, S., Brands, I., & Jelicic, M. (2010). Assessing trait memory distrust: Psychometric properties of the Squire Subjective Memory Questionnaire. *Legal and Criminological Psychology*, 15, 373-384. doi:10.1348/135532509X471960
- van Bergen, S., Horselenberg, R., Merckelbach, H., Jelicic, M., & Beckers, R. (2010). Memory distrust and acceptance of misinformation. *Applied Cognitive Psychology*, 24, 885-896. doi: 10.1002/acp.1595
- van der Werf, S. P., & Vos, S. H. (2011). Memory worries and self-reported daily forgetfulness: A psychometric evaluation of the Dutch translation of the multifactorial memory questionnaire. *The Clinical Neuropsychologist*, 25, 244-268. doi: 10.1080/13854046.2010.543290
- Van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior*, 21, 661–685. doi: 10.1023/A:1024812831576
- Ventura, P., Livingston, L. A., & Shah, P. (2018). Adults have moderate-to-good insight into their face recognition ability: Further validation of the 20-item Prosopagnosia Index in a Portuguese sample. *Quarterly Journal of Experimental Psychology*, 71, 2677-2679. doi: 10.1177/1747021818765652
- Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. *Applied Cognitive Psychology*, 20(1), 17-31. doi: 10.1002/acp.1166
- Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of

- confidence. *Journal of Experimental Psychology: Applied*, 14, 50-60. doi: 10.1037/1076-898X.14.1.50
- Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior*, 36, 28–36. doi:10.1037/h0093976
- Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12-second rule. *Journal of Experimental Psychology: Applied*, 10, 139–147. doi:10.1037/1076-898X.10.3.139
- Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology*, 83, 360-376. doi: 10.1037/0021-9010.83.3.360
- Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the postidentification-feedback effect be moderated? *Psychological Science*, 10, 138–144. doi: 10.1111/1467-9280.00121
- Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, 54, 277-295. doi: 10.1146/annurev.psych.54.101601.145028
- Wells, G. L., Olson, E. A., & Charman, S. D. (2002). The confidence of eyewitnesses in their identifications from lineups. *Current Directions in Psychological Science*, 11, 151–154. doi: 10.1111/1467-8721.00189
- Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, 22, 603-647. doi: 10.1023/A:1025750605807
- Wheatcroft, J. M., Wagstaff, G. F., & Manarin, B. (2015). The influence of delay and item difficulty on eyewitness confidence and accuracy. *International Journal of Humanities and Social Science Research*, 1, 1-9. ISSN: 2371-1655

References

- Windschitl, P. D., & Chambers, J. R. (2004). The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 30, 198–215. doi: 10.1037/0278-7393.30.1.198
- Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, 18, 10–65. doi:10.1177/152910061668696
- Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the reliability of eyewitness memory. *Perspectives on Psychological Science*, 13, 324–335. doi: 10.1177/1745691617734878
- Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L., 3rd. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *The American Psychologist*, 70, 515–526. doi: 10.1037/a0039510
- Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron*, 44, 889–898. doi: 10.1016/j.neuron.2004.11.018
- Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., ... Dong, Q. (2010). Individual differences in false memory from misinformation: Cognitive factors. *Memory*, 18, 543–555. doi: 10.1080/09658211.2010.487051
- Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., ... Dong, Q. (2010). Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Personality and Individual Differences*, 48, 889–894. doi: 10.1016/j.paid.2010.02.016

Appendices

APPENDIX 1

Chapters 2 and 6

Cued-Recall questions about the video

Now we would like to ask you some questions about the video you just watched. To answer the questions, consider that the name of the woman in the video is Lisa. We will also ask you to give a confidence rating for each one of your answers, in a scale that ranges from 0% (**‘not at all certain’**) and then 10, 20, 30, ... to 100% (**‘totally certain’**). If you do not know the answer for a question write something and assign your confidence rating as 0.

1. What did Lisa do in the room with the sofa? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
2. What was the colour of the sofa? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
3. What was the colour of Lisa's hoodie/sweatshirt? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
4. What was the first vegetable Lisa chops in the kitchen? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
5. What brand or make is Lisa's hoodie/sweatshirt? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
6. What did Lisa do in the room with a view of the garden? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
7. How many pans did Lisa use to cook her food? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
8. What room was the washing machine in? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
9. What kind of footwear was Lisa wearing? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
10. How many cushions were there on the sofa? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
11. What did Lisa do in the kitchen before cooking? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
12. What colour was Lisa's trousers? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
13. One of the cushions on the sofa was a different shape to the others; which shape was it? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

Appendices

14. What did Lisa do in the TV room? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

15. How would you describe Lisa's hairstyle? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

16. What colour was the phone in the room with the sofa? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

17. In the room with a view into the garden, what object was hanging on a chair? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

APPENDIX 2

Chapters 2 and 6

General Knowledge Questions

Now we will ask you to answer some general knowledge questions. Please, try to complete this task in the best way you can. There is no problem if you do not know all the answers, but try your best. We will also ask you to give a confidence rating for each one of your answers, in a scale that ranges from 0% (**'not at all certain'**) and then 10, 20, 30, ... to 100% (**'totally certain'**). If you do not know the answer for a question write something and assign your confidence rating as 0.

1. What kind of celestial body is the moon? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
2. Who discovered the magnetic field of electric current? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
3. In what country is the Yellow Stone National Park? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
4. What was the first metal used by the man? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
5. What is the world largest desert? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
6. In what country is the mount Everest located? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
7. What is the device used for measuring altitudes? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
8. What country is known as the Land of Rising Sun? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
9. Deficiency of Iron in the blood leads to...? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
10. Which is the hottest planet in the solar system? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
11. Which gas is used for the preparation of Soda water? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
12. Which country lost the largest number of people in the 2nd world war? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------
13. What type of animal is Bambi? _____

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

Appendices

14. What part of the body produces insulin? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
15. Who was the Greek or Roman God of War? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
16. What is the religion of the Dalai Lama? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
17. What is the name of the poker hand containing three of a kind and a pair? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
18. Tony Awards are presented in which field of the arts? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
19. How many legs does an ant have? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
20. Which gas in our atmosphere protects Earth from ultra violet radiation? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
21. In which country were the 2008 Olympic Games held? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
22. What type of music did Louis Armstrong play? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
23. What do we call the study of birds? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
24. Who was the only person to win a Nobel Prize in two different sciences? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
25. Who killed Goliath? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
26. Which is the longest river on the earth? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
27. Who was the first person to walk on the moon? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
28. What is the most populous country in the world? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
29. What was the name of the town which was the birthplace of Jesus Christ? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
30. Cirrus or cumulus are examples of...? _____
- | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|

APPENDIX 3

Chapter 4

Table S4.1

Model Fit Indices for Fixed Intercept Models (Model 1) and Random Intercept Models (Model 2) for Each Condition

	Biased				Unbiased			
	Choosers		Nonchoosers		Choosers		Nonchoosers	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Log likelihood	-248.54	-247.92	-196.96	-196.96	-265.86	-265.40	-196.06	-196.06
AIC	513.08	515.84	409.92	413.92	547.72	550.80	408.11	412.11
BIC	545.29	556.09	440.46	452.09	579.67	590.74	438.67	450.31
ICC Participant	-	0.11	-	0.01	-	0.08	-	0.00
ICC Lineup	-	.00	-	0.00	-	0.00	-	0.00

APPENDIX 4

Chapter 5

Table S5.1

Model fit indices for fixed intercept models (Model 1) and random intercept models (Model 2) for each outcome variable

	Correct Identifications		Correct rejections		False identifications		Confidence	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Log likelihood	-469.69	-457.92	-527.47	-500.66	-882.83	-845.77	-3602.0	-3504.7
AIC	953.38	933.83	1068.9	1019.3	1779.7	1709.5	7219.9	7029.3
BIC	986.28	976.13	1101.8	1061.6	1817.4	1758.1	7263.1	7083.2
ICC Participant	-	.015	-	.005	-	.050	-	.185
ICC Lineup	-	.081	-	.119	-	.097	-	.060

Table S5.2

Logistic multilevel models testing metamemory scores as predictors of choosing (0 = non-chooser, 1 = chooser)

	<i>OR</i>	95% CI	<i>p</i>
Fixed coefficients			
EMS-Contentment	0.98	0.87, 1.09	.73
EMS-Discontentment	1.00	0.90, 1.12	.88
EMS-Strategies	1.16	1.04, 1.30	.005**
MMQ-Contentment	0.98	0.86, 1.12	.80
MMQ-Ability	0.89	0.78, 1.02	.12
MMQ-Strategy	0.92	0.82, 1.04	.20
Random parameters	Variance		<i>SD</i>
Level 1: Lineup	.22		.47

Note. ** indicates $p < .01$.

APPENDIX 5

Chapter 6

Free Recall Test

In the space provided, report all the details that you can remember about the video, including the sequence of actions and events, and the people that were involved. If you recall information or specific details out of the order in which they happened, report these details as they come to mind (i.e., do not leave out any details.) Do not guess about details that you cannot remember. Feel free to use full sentences or bullet points – but please make sure your report is as complete and accurate as possible.

Please write your report using only the white boxes (**do not write in the grey lines**), the grey lines will be used later.

1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	
11.	
12.	
13.	
14.	
15.	
16.	
17.	
18.	
19.	

Appendices

20.	
21.	
22.	
23.	
24.	
25.	
26.	
27.	
28.	
29.	
30.	
31.	
32.	
33.	
34.	
35.	
36.	
37.	
38.	
39.	
40.	
41.	
42.	
43.	
44.	
45.	
46.	
47.	
48.	
49.	

50.	
-----	--

APPENDIX 6

Table S6.1

Multiple linear regression models including metamemory scales as predictors of eyewitness free recall performance

Predictors	<i>B</i> (<i>SE</i>)	<i>B</i> CI	β	<i>t</i>	<i>p</i>
Free recall accuracy					

Appendices

MMQ-Contentment	1.04 (<0.001)	[-0.003, 0.003]	<0.01	-0.05	.96
MMQ-Ability	<0.001 (<0.001)	[-0.005, 0.002]	-0.08	-0.82	.82
MMQ-Strategies	<0.001 (<0.001)	[-0.002, 0.003]	0.02	0.22	.41
SSMQ	<0.001 (<0.001)	[-0.002, 0.001]	-0.08	-0.85	.39
General Memory Ability	<0.001 (<0.001)	[-0.001, 0.001]	0.01	0.11	.21
Face Memory Ability	<0.001 (<0.001)	[-0.001, 0.001]	0.12	1.27	.91
Free recall completeness					
MMQ-Contentment	0.66 (1.69)	[-2.68, 4.01]	0.04	0.39	.69
MMQ-Ability	-0.78 (2.06)	[-1.94, 4.26]	-0.04	-0.37	.70
MMQ-Strategies	1.15 (1.57)	[-4.85, 3.28]	0.06	0.73	.46
SSMQ	0.23 (0.97)	[-1.68, 2.14]	0.02	0.24	.81
General Memory Ability	-0.12 (0.70)	[-0.89, 1.23]	-0.02	-0.17	.86
Face Memory Ability	0.16 (0.54)	[-1.50, 1.26]	0.03	0.31	.76
Free recall confidence					
MMQ-Contentment	-0.04 (0.55)	[-0.21, 0.12]	-0.04	-0.49	.62
MMQ-Ability	0.13 (0.10)	[-0.07, 0.34]	0.12	1.25	.21
MMQ-Strategies	-0.02 (0.08)	[-0.18, 0.13]	-0.02	-0.31	.76
SSMQ	0.01 (0.05)	[-0.08, 0.11]	0.03	0.36	.72
General Memory Ability	0.02 (0.03)	[-0.05, 0.09]	0.05	0.54	.59
Face Memory Ability	0.04 (0.03)	[-0.01, 0.09]	0.12	1.31	.19
Free recall over/underconfidence					
MMQ-Contentment	-0.008 (0.064)	[-0.024, 0.017]	-0.033	-0.34	.72
MMQ-Ability	-0.003 (0.010)	[-0.042, 0.007]	-0.140	-1.36	.17
MMQ-Strategies	-0.001 (0.009)	[-0.019, 0.018]	-0.003	-0.35	.72
SSMQ	0.004 (0.012)	[-0.007, 0.016]	0.069	0.75	.45
General Memory Ability	0.003 (0.004)	[-0.005, 0.011]	0.075	0.72	.47
Face Memory Ability	-0.002 (0.003)	[-0.008, 0.004]	-0.056	-0.62	.53

APPENDIX 7

Ethics approval (Chapters 2 to 6)

Each of the experiments presented in this thesis received ethical approval from the University of Portsmouth's Science Faculty Ethics Committee (SFEC). Below, the letter of favourable opinion is attached for each of the experiments reported in this thesis.



Mr Renan Saraiva
Department of Psychology
University of Portsmouth

Renan.Saraiva@port.ac.uk

Science Faculty Ethics Committee

Science Faculty Office
University of Portsmouth
St Michael's Building
White Swan Road
PORTSMOUTH
PO1 2DT

023 9284 3379

ethics-sci@port.ac.uk

27 March 2019

FAVOURABLE ETHICAL OPINION – NOTIFICATION OF SUBSTANTIAL AMENDMENT

Study Title: The role of metamemory and confidence-accuracy relationship during eyewitness identification tasks

Reference Number: SFEC 2017-028A

Date Submitted: 12 March 2019

Thank you for submitting your proposal amendment to the Science Faculty Ethics Committee (SFEC) for ethical review in accordance with current procedures.

I am pleased to inform you that SFEC was content to grant a favourable ethical opinion of this proposal amendment on the basis described in the submitted documents listed at Annex A, and subject to standard general conditions (*See Annex B*).

Please note that the favourable opinion of SFEC does not grant permission or approval to undertake the research. Management permission or approval must be obtained from any host organisation, including the University of Portsmouth or supervisor, prior to the start of the study.

Wishing you every success in your research

A handwritten signature in black ink, appearing to read 'John C', enclosed in a rectangular box.

Dr John Crossland
Vice Chair Science Faculty Ethics Committee

Annexes

A - Documents reviewed

B - After ethical review - Guidance for researchers



Renan Saraiva
Department of Psychology
University of Portsmouth

Renan.Saraiva@port.ac.uk

Science Faculty Ethics Committee

Science Faculty Office
University of Portsmouth
St Michael's Building
White Swan Road
PORTSMOUTH
PO1 2DT

023 9284 3379
ethics-sci@port.ac.uk

2 November 2018

FAVOURABLE ETHICAL OPINION – FOLLOWING RESUBMISSION

Study Title: Using metamemory measures to estimate eyewitness over / under confidence in repeated identifications

Reference Number: SFEC 2018-087

Date Resubmitted: 29 October 2018

Thank you for resubmitting your application to the Science Faculty Ethics Committee (SFEC) for ethical review in accordance with current procedures, for making the requested changes following the first SFEC review, and for the clarifications provided.

I am pleased to inform you that SFEC was content to grant a favourable ethical opinion of the above research on the basis described in the submitted documents listed at Annex A, and subject to standard general conditions (*See Annex B*).

Please note that the favourable opinion of SFEC does not grant permission or approval to undertake the research. Management permission or approval must be obtained from any host organisation, including the University of Portsmouth or supervisor, prior to the start of the study.

Wishing you every success in your research

A handwritten signature in black ink, appearing to read 'Jim House'.

Dr Jim House
Chair, Science Faculty Ethics Committee

Annexes

A - Documents reviewed

B - After ethical review - Guidance for researchers

Information:

Professor Lorraine Hope - PhD Supervisor
Rose Barrand - Faculty Administrator



Mr Renan Saraiva
Department of Psychology
University of Portsmouth

Renan.Saraiva@port.ac.uk

Science Faculty Ethics Committee

Science Faculty Office
University of Portsmouth
St Michael's Building
White Swan Road
PORTSMOUTH
PO1 2DT

023 9284 3379
ethics-sci@port.ac.uk

30 March 2017

FAVOURABLE ETHICAL OPINION – FOLLOWING RESUBMISSION

Study Title: The role of metamemory and confidence-accuracy relationship during eyewitness identification tasks

Reference Number: SFEC 2017-028

Date Resubmitted: 27 March 2017

Thank you for resubmitting your application to the Science Faculty Ethics Committee (SEFC) for ethical review in accordance with current procedures, for making the requested changes following the first SFEC review, and for the clarifications provided.

I am pleased to inform you that SFEC was content to grant a favourable ethical opinion of the above research on the basis described in the submitted documents listed at Annex A, and subject to standard general conditions (*See Annex B*).

Please note that the favourable opinion of SFEC does not grant permission or approval to undertake the research. Management permission or approval must be obtained from any host organisation, including the University of Portsmouth or supervisor, prior to the start of the study.

Wishing you every success in your research

A handwritten signature in black ink, appearing to read 'John Crossland'.

Dr John Crossland
Vice Chair, Science Faculty Ethics Committee

Annexes

- A - Documents reviewed
- B - After ethical review - Guidance for researchers

Information:

Dr James Ost - Co-Investigator
Professor Loraine Hope - PhD Supervisor



UNIVERSITY OF
PORTSMOUTH

Renan Saraiva
Department of Psychology
University of Portsmouth

Renan.Saraiva@port.ac.uk

Science Faculty Ethics Committee

Science Faculty Office
University of Portsmouth
St Michael's Building
White Swan Road
PORTSMOUTH
PO1 2DT

023 9284 3379
ethics-sci@port.ac.uk

23 April 2018

FAVOURABLE ETHICAL OPINION – WITH CONDITIONS

Study Title: The Role of Eyewitness-Specific Metamemory in Eyewitness Identification Performance for Biased and Unbiased Lineups

Reference Number: SFEC 2018-040

Date Submitted: 05 April 2018

Thank you for submitting your application to the Science Faculty Ethics Committee (SFEC) for ethical review in accordance with current procedures.

I am pleased to inform you that SFEC was content to grant a favourable ethical opinion of the above research on the basis described in the submitted documents listed at Annex A, and subject to standard general conditions (*See Annex B*), and the following specific minor conditions:

Condition¹

A. Could the PI supply the stimulus video or a link to the video?

Advisory Notes

These advisory notes are given in good faith and it is hoped they are accepted as such. You do not need to adhere to these comments, or respond to them, unless you wish to.

i. Are 400 participants really necessary? Could adequate power be attained with fewer participants?

ii. The PI may wish to consider sending PIS 24 hours before the study.

iii. The rewards for taking part seem a little unbalanced - the PI may wish to consider making the reward more consistent.

¹ The favourable opinion given is dependent upon the study adhering to the conditions stated, which are based on the application document(s) submitted. It is appreciated that Principal Investigators may wish to challenge conditions or propose amendments to these. In that case, please consider the favourable opinion *suspended*, and simply make your case for amending or discarding conditions in writing as you would an application resubmission following ethical review.



Renan Benigno Saraiva
Department of Psychology
University of Portsmouth

renan.saraiva@port.ac.uk

Science Faculty Ethics Committee

Science Faculty Office
University of Portsmouth
St Michael's Building
White Swan Road
PORTSMOUTH
PO1 2DT

T: 023 9284 3379
ethics-sci@port.ac.uk

14 December 2016

FAVOURABLE ETHICAL OPINION

Study Title: The role of metamemory on the identification of reliable eyewitnesses

Reference Number: SFEC 2016-110

Date Resubmitted: 12 December 2016

Thank you for resubmitting your application to the Science Faculty Ethics Committee (SEFC) for ethical review in accordance with current procedures¹, and for making the requested changes following the first SFEC review, and for the clarifications provided.

I am pleased to inform you that SFEC was content to grant a favourable ethical opinion of the above research on the basis described in the submitted documents listed at Annex A, and subject to standard general conditions (*See Annex B*).

Please note that the favourable opinion of SFEC does not grant permission or approval to undertake the research. Management permission or approval must be obtained from any host organisation, including the University of Portsmouth or supervisor, prior to the start of the study.

Wishing you every success in your research

A handwritten signature in black ink, appearing to read 'John Crossland'.

Dr John Crossland
Vice Chair, Science Faculty Ethics Committee

Annexes

- A - Documents reviewed
- B - After ethical review - Guidance for researchers

Information:

Prof Lorraine Hope – First supervisor

¹ Procedures for Ethical Review, Science Faculty Ethics Committee, University of Portsmouth, October 2012 (to be updated).

FORM UPR16

Research Ethics Review Checklist

Please include this completed form as an appendix to your thesis (see the Research Degrees Operational Handbook for more information)



Postgraduate Research Student (PGRS) Information		Student ID:	840532
PGRS Name:	Renan Saraiva		
Department:	Psychology	First Supervisor:	Lorraine Hope
Start Date: (or progression date for Prof Doc students)	01/09/2016		
Study Mode and Route:	Part-time <input type="checkbox"/> Full-time <input checked="" type="checkbox"/>	MPhil <input type="checkbox"/> PhD <input checked="" type="checkbox"/>	MD <input type="checkbox"/> Professional Doctorate <input type="checkbox"/>
Title of Thesis:	The role of metamemory in eyewitness performance		
Thesis Word Count: (excluding ancillary data)	41.865		

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:

(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: <http://www.ukrio.org/what-we-do/code-of-practice-for-research/>)

a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>
b) Have all contributions to knowledge been acknowledged?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>
e) Does your research comply with all legal, ethical, and contractual requirements?	YES <input checked="" type="checkbox"/> NO <input type="checkbox"/>

Candidate Statement:

I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)

Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):

SFEC 2017-028A
SFEC 2018-087
SFEC 2017-028
SFEC 2018-040
SFEC 2016-110

If you have *not* submitted your work for ethical review, and/or you have answered 'No' to one or more of questions a) to e), please explain below why this is so:

Signed (PGRS):

Renan Benigno Saraiva

Date: 02/08/2019